

## 10. Stochastic Processes

### Basic Probability

Probability theory is a large subject that has developed deep philosophical roots and corresponding firmly held differences of opinion about its interpretation. Some of these disputes are very important, but for our purposes we will think of a probability density as a frequency function (histogram) in the limit of an infinite number of trials of an experiment as the bin-sizes go to zero. We assume that such a limit exists, and can be written for a random variable  $y$ , as  $p_y(Y)$ , where the subscript is the variable, and the argument  $Y$  is the *values* that  $y$  can take on (distinguishing the physical variable, e.g. a temperature, from its numerical value). Sometimes the subscript is dropped when the meaning is otherwise clear.

Continued on next page...

The use of  $p_y(Y)$  in practice means that one is referring to the probability that  $y$  lies in the interval  $Y \leq y \leq Y + dY$  as  $p_y(Y) dY$ , that is, lying in some differential interval.

Two very useful probability densities are the uniform one over interval  $L$

$$\begin{aligned} p_y(Y) &= \frac{1}{L}, -\frac{L}{2} \leq Y \leq \frac{L}{2} \\ &= \Pi(Y/L) \end{aligned} \quad (10.1)$$

and the normal (or  $\Sigma$ ),

$$p_y(Y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(Y-m)^2/(2\sigma^2)}. \quad (10.2)$$

The latter is also written  $G(m, \sigma^2)$  to denote a normal (Gaussian) density with mean  $m$ , and variance  $\sigma^2$ . Both (10.1,10.2) satisfy the necessary requirements,  $p \geq 0$ , and

$$\int_{-\infty}^{\infty} p_y(Y) dY = 1. \quad (10.3)$$

Define a bracket,  $\langle . \rangle$  as the averaging operator with meaning of the integral over all possible values of the argument times the probability density. Thus,

$$\langle y \rangle = \int_{-\infty}^{\infty} Y p_y(Y) dY = m \quad (10.4)$$

(the mean, or center of mass),

$$\langle y^2 \rangle = \int_{-\infty}^{\infty} Y^2 p_y(Y) dY \quad (10.5)$$

the second moment,

$$\langle (y - \langle y \rangle)^2 \rangle = \int_{-\infty}^{\infty} (Y - \langle y \rangle)^2 p_y(Y) dY = \sigma_y^2 \quad (10.6)$$

the variance (second moment about the mean), etc. An important simple result is that if  $a$  is constant (not random), then

$$\langle ay \rangle = a \langle y \rangle. \quad (10.7)$$

Let  $f(y)$  be any function of random variable,  $y$ . It follows from the frequency function definition of the probability density, that

$$\langle f(y) \rangle = \int_{-\infty}^{\infty} f(Y) p_y(Y) dY. \quad (10.8)$$

Eq. (10.7) is evidently a special case. Often it is useful to find the probability density of  $f$  from that of  $y$ . We suppose that the function is invertible so that  $y(f)$  is known. Then the line segment  $dY$  is mapped into a line-segment  $dF$ , by the rule,

$$dY = \frac{dy(F)}{dF} dF, \quad (10.9)$$

we have immediately,

$$p_g(F) dF = p_y(Y(F)) \frac{dy(F)}{dF} dF. \quad (10.10)$$

A special case would be  $f = ay + b$ , or  $y = (f - b)/a$ ,  $dy(F)/dF = 1/a$ , and thus

$$p_g(F) = p_y((F - b)/a) / a. \quad (10.11)$$

If this result is applied to (10.2), we have

$$p_g(F) = \frac{1}{\sqrt{2\pi}a\sigma} e^{-(F-b-ma)^2/(2a^2\sigma^2)} \quad (10.12)$$

which is  $G(b - ma, a^2\sigma^2)$ . Note that by choosing  $b = m/\sigma$ ,  $a = 1/\sigma$ , that the new probability density is  $G(0, 1)$ —a standard form that is the one usually tabulated.

If the derivative  $dy/dF$  should be zero or infinity, it implies that the mapping from  $y$  to  $f$ , or from  $f$  to  $y$  is not unique, and some care is needed (the differentials are not uniquely mapped). So consider  $G(0, 1)$ ,

$$p_y(Y) = \frac{1}{\sqrt{2\pi}} e^{-Y^2/2} \quad (10.13)$$

and suppose  $\xi = y^2$ . Then clearly the probability that  $\xi$  is less than zero is zero, and both  $\pm y$  map onto the same value of  $\xi$  indicated by,

$$\frac{dy}{d\xi} = \frac{1}{2y} = \frac{1}{2\sqrt{\xi}} \quad (10.14)$$

becoming infinite at  $\xi = 0$ . We can deduce therefore that,

$$p_\xi(X) = \begin{cases} \frac{1}{\sqrt{2\pi\sqrt{X}}} e^{-X/2}, & X \geq 0 \\ 0, & X < 0 \end{cases} \quad (10.15)$$

multiplying by 2 to account for the negative  $Y$  contributions, too. Probability density (10.15) is known as “chi-square with one degree-of-freedom” usually written  $\chi_1^2$ . “Degrees-of-freedom” will be defined later. For future reference, note that if  $\langle y^2 \rangle = \sigma^2$ , still with zero mean, then Eq. (10.15) becomes

$$p_\xi(X) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi\sqrt{X}}} e^{-X/(2\sigma^2)}, & X \geq 0 \\ 0, & X < 0 \end{cases} \quad (10.16)$$

It can become confusing to keep track of mapping functions  $g(y)$  which are not unique, and a more systematic approach than used to find (10.15) is desirable. Introduce the “probability distribution function”,

$$P_y(Y) = \int_{-\infty}^Y p_y(Y') dY', \quad (10.17)$$

which has the properties,  $dP_y/dY \geq 0$ ,  $P_y(\infty) = 1$ ,  $dP_y/dY = p_y(Y)$ . The interpretation of  $P_y(Y)$  is as the probability that  $y$  is less than or equal to  $Y$ .

Then for the above case, with  $\xi = y^2$  and  $y$  being Gaussian,

$$P_\xi(X) = \text{probability that } \{-\sqrt{X} \leq Y \leq \sqrt{X}\} = P_y(\sqrt{X}) - P_y(-\sqrt{X}) \quad (10.18)$$

$$= \int_{-\infty}^{\sqrt{X}} \frac{1}{\sqrt{2\pi}} e^{-Y^2/2} dY - \int_{-\infty}^{-\sqrt{X}} \frac{1}{\sqrt{2\pi}} e^{-Y^2/2} dY. \quad (10.19)$$

And,

$$p_\xi(X) = \frac{d}{dX} P_\xi(X) = \frac{d}{dX} \left[ \int_{-\infty}^{\sqrt{X}} \frac{1}{\sqrt{2\pi}} e^{-Y^2/2} dY - \int_{-\infty}^{-\sqrt{X}} \frac{1}{\sqrt{2\pi}} e^{-Y^2/2} dY \right] \quad (10.20)$$

$$= \left[ \frac{1}{\sqrt{2\pi}\sqrt{X}} e^{-\frac{1}{2}X} \right], \quad X \geq 0, \quad (10.21)$$

identical to (10.15). “Leibniz’s rule” for differentiation of a variable upper bound of integration was used. This approach is quite general, as long as expressions such as (10.18) can be constructed.

If there are two or more random variables,  $\xi_i, i = 1, 2, \dots, m$  we can discuss their “joint” or “multivariate probability densities”,  $p_{\xi_1 \xi_2 \dots}(\Xi_1, \Xi_2, \dots, \Xi_m)$ ; these are to be thought of as derived from the limits of a counting experiment in which  $\xi_1, \xi_2, \dots$  are measured many times, and then binned by the values observed. The limit, as the number of such observations goes to infinity and as the bin size goes to zero, is supposed to exist.

If a joint probability density factors,

$$p_{\xi_1 \xi_2 \dots}(\Xi_1, \Xi_2, \dots, \Xi_m) = p_{\xi_1}(\Xi_1) p_{\xi_2}(\Xi_2) \dots p_{\xi_m}(\Xi_m) \quad (10.22)$$

then the  $\xi_i$  are said to be independent.

*Example.* The general,  $m$ -dimensional joint Gaussian probability density is defined as,

$$p_{\xi_1 \xi_2 \dots}(\Xi_1, \Xi_2, \dots, \Xi_m) = \frac{1}{(2\pi)^{m/2} |\mathbf{R}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\xi} - \mathbf{m})^T \mathbf{R}^{-1} (\boldsymbol{\xi} - \mathbf{m})\right). \quad (10.23)$$

Here  $\boldsymbol{\xi} = [\xi_1, \dots, \xi_m]^T$ ,

$$\mathbf{R} = \left\{ \begin{array}{cccc} \langle (\xi_1 - m_1)(\xi_1 - m_1) \rangle & \langle (\xi_1 - m_1)(\xi_2 - m_2) \rangle & \dots & \langle (\xi_1 - m_1)(\xi_m - m_m) \rangle \\ \langle (\xi_2 - m_2)(\xi_1 - m_1) \rangle & \langle (\xi_2 - m_2)(\xi_2 - m_2) \rangle & \dots & \langle (\xi_2 - m_2)(\xi_m - m_m) \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle (\xi_m - m_m)(\xi_1 - m_1) \rangle & \langle (\xi_m - m_m)(\xi_2 - m_2) \rangle & \dots & \langle (\xi_m - m_m)(\xi_m - m_m) \rangle \end{array} \right\} \quad (10.24)$$

and  $|\mathbf{R}|$  is the determinant.  $\mathbf{R}$  can be written in a variety of ways including

$$\mathbf{R} = \left\{ \begin{array}{cccc} \sigma_1^2 & \sigma_1 \sigma_2 \rho_{12} & \dots & \sigma_1 \sigma_m \rho_{1m} \\ \sigma_2 \sigma_1 \rho_{21} & \sigma_2^2 & \dots & \sigma_2 \sigma_m \rho_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_m \sigma_1 \rho_{m1} & \sigma_m \sigma_2 \rho_{m2} & \dots & \sigma_m^2 \end{array} \right\} \quad (10.25)$$

where the  $\sigma_i^2$  are the corresponding variances about the mean, and the  $\rho_{ij} \equiv \langle (\xi_i - m_i)(\xi_j - m_j) \rangle / \sigma_i \sigma_j = \rho_{ji}$  are called correlations (discussed below).

The important special case of two normal variables (“bivariate normal”) can be written as

$$p_{\xi_1\xi_2}(\Xi_1, \Xi_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left( \frac{(\Xi_1 - m_1)^2}{\sigma_1^2} - \frac{2\rho(\Xi_1 - m_1)(\Xi_2 - m_2)}{\sigma_1\sigma_2} + \frac{(\Xi_2 - m_2)^2}{\sigma_2^2} \right) \right\}, \quad (10.26)$$

where  $\rho$  is defined as  $\rho = \langle (\xi_1 - m_1)(\xi_2 - m_2) \rangle / \sigma_1\sigma_2$ , and taken up immediately below. Other variants of (10.26) are possible.

*Exercise.* Show that if all  $\rho_{ij} = 0, i \neq j$ , that (10.23) reduces to the product of  $m$ -univariate normal distributions, hence showing that uncorrelated normal variates are also independent.

The joint expectations about the means (moments about the means) are, for two variates,

$$\langle \xi'_1 \xi'_2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\Xi_1 - m_1)(\Xi_2 - m_2) p_{\xi_1\xi_2}(\Xi_1, \Xi_2) d\Xi_1 d\Xi_2. \quad (10.27)$$

We will use  $\xi'_i$  to denote the variable with its mean removed. If  $\xi'_1, \xi'_2$  are independent, it follows that  $\langle \xi'_1 \xi'_2 \rangle = \langle \xi'_1 \rangle \langle \xi'_2 \rangle = 0$ .

If  $\langle \xi'_1 \xi'_2 \rangle \neq 0$ , (suppressing the prime), then they are said to be “correlated” variables. This implies that a knowledge of one of them provides some predictive capacity for the other. We can use the idea of conditional probability, e.g., the probability that  $\xi_2$  takes on a particular value (or range of values)  $\Xi_2$  given that  $\xi_1 = \Xi_1$ , which we write as

$$p_{\xi_2|\xi_1}(\Xi_2|\Xi_1). \quad (10.28)$$

The textbooks all show that if  $\xi_1, \xi_2$  are independent  $p_{\xi_2|\xi_1}(\Xi_2|\Xi_1) = p_{\xi_2}(\Xi_2)$ , that is, knowledge of the value of  $\xi_1$  then contains no predictive information about  $\xi_2$ . More generally, suppose that we try to predict  $\xi_2$  from  $\xi_1$  in the form

$$\xi_2 = a\xi_1 + \varepsilon \quad (10.29)$$

where by definition  $\varepsilon$  is independent of  $\xi_1$ . Forming  $\langle \xi_2 \xi_1 \rangle = a \langle \xi_1^2 \rangle + \langle \varepsilon \xi_1 \rangle = a \langle \xi_1^2 \rangle$ , or

$$a = \frac{\langle \xi_2 \xi_1 \rangle}{\langle \xi_1^2 \rangle}, \quad (10.30)$$

which would vanish if  $\langle \xi_1 \xi_2 \rangle = 0$ . Define the “correlation coefficient”

$$\rho = \frac{\langle \xi_2 \xi_1 \rangle}{\langle \xi_1^2 \rangle^{1/2} \langle \xi_2^2 \rangle^{1/2}}. \quad (10.31)$$

It is straightforward to show that  $|\rho| \leq 1$ . (e.g., Priestley, p. 79) We have,

$$a = \rho \frac{\langle \xi_2^2 \rangle^{1/2}}{\langle \xi_1^2 \rangle^{1/2}}, \quad (10.32)$$

and it follows that the fraction of the variance in  $\xi_2$  which is correlated with (predictable by knowledge of  $\xi_1$ ) is just,

$$\langle \xi_2^2 \rangle \rho^2 \quad (10.33)$$

(the “correlated power”), and the part of the variance which is not predictable, is

$$\langle \xi_2^2 \rangle > (1 - \rho^2) \quad (10.34)$$

which is the “uncorrelated power”. If  $\xi_1, \xi_2$  are independent, they are uncorrelated; if they are uncorrelated, they need not be independent.

Let there be  $m$  random variables  $\xi_1, \dots, \xi_m$ , and suppose we define  $m$  new variables  $\eta_1, \dots, \eta_m$  which are functions of the original variables. Conservation of area rules lead to the conclusion that if the joint probability density of the  $\xi_i$  is  $p_{\xi_1 \xi_2 \dots \xi_m}(\Xi_1, \Xi_2, \dots, \Xi_m)$ , then the joint probability density for the  $\eta_i$  is

$$p_{\eta_1 \eta_2 \dots}(\eta_1, \eta_2, \dots, \eta_m) = \quad (10.35)$$

$$p_{\xi_1 \xi_2 \dots \xi_m}(\Xi_1(\eta_1, \dots, \eta_m), \Xi_2(\eta_1, \dots, \eta_m), \dots, \Xi_m(\eta_1, \dots, \eta_m)) \frac{\partial(\Xi_1, \Xi_2, \dots, \Xi_m)}{\partial(\eta_1, \dots, \eta_m)}$$

where

$$\frac{\partial(\Xi_1, \Xi_2, \dots, \Xi_m)}{\partial(\eta_1, \dots, \eta_m)} \quad (10.36)$$

is the Jacobian of the transformation between the two sets of variables.

In one dimension, the most common and useful transformations are of the form  $\eta = a\xi + b$ ,  $a, b$  constant. In most cases, we work with canonical probability densities, such that e.g., the mean is zero, and the variance unity. The linear transformation, with Jacobian  $a$  or  $a^{-1}$  permits one to use these standard densities for random variables with arbitrary means and variances (see Eq. 10.12). All the ideas concerning correlation, predictability and independence are generalizable to more than two variables, through multiple and partial correlation studies, but these are left to the references.

*Example.* Let  $x, y$  be two uncorrelated (and hence independent) Gaussian random variables of zero mean and variance  $\sigma^2$ . Define  $r = \sqrt{x^2 + y^2}$  and  $\phi = \tan^{-1}(y/x)$ . We seek the joint probability density for  $r, \phi$ , and their univariate probability densities. The joint probability density function is

$$p_{x,y}(X, Y) = \frac{1}{2\pi\sigma} \exp[-(X^2 + Y^2)/2\sigma^2].$$

The Jacobian of the transformation from  $(x, y)$  to  $(r, \phi)$  is just  $r$  (Cartesian to polar coordinates). Then the joint probability density of the new variables is

$$p_{r,\phi}(R, \Phi) = \frac{R}{2\pi\sigma} \exp[-R^2/2\sigma^2]$$

Integrating out the  $\Phi$  variable over its entire range,  $-\pi \leq \phi \leq \pi$ , ( $\Phi$  doesn't actually appear), we have immediately,

$$p_r(R) = \frac{R}{\sigma} \exp(-R^2/2\sigma^2), \quad (10.37)$$

and by inspection, it must be true that

$$p_\phi(\Phi) = \frac{1}{2\pi}.$$

Thus the phase has a uniform distribution  $-\pi \leq \phi \leq \pi$ , and the amplitude and phase are uncorrelated with each other. Note that the probability density for  $r$  is called the “Rayleigh distribution”.

*Exercise.* Find the mean and variance of a Rayleigh distributed variable.

For time series work, the most important  $m$ -dimensional probability density is the Gaussian or normal one (10.23). As the textbooks all show, the normal probability density has several important special properties. One of them is that it is completely specified by its mean,  $\mathbf{m}$ , and covariance matrix,  $\mathbf{R} = \{ \langle (\xi_i - m_i) (\xi_j - m_j) \rangle \}$ , and that if any pair is uncorrelated, they are also independent.

### Adding Independent Variates. Characteristic Functions

A common problem is to determine the probability density of a sum of variates, e.g.,  $\xi = \xi_1 + \xi_2 + \dots$ . Obtaining the probability density is best done by undertaking a seeming digression. Consider an arbitrary random variable  $y$  with known probability density  $p_y(Y)$ , and the function  $g(y) = e^{iyt}$ . By (10.8), its expected value is

$$\phi_y(t) = \langle e^{iyt} \rangle = \int_{-\infty}^{\infty} p_y(Y) e^{iYt} dY, \quad (10.38)$$

that is, up to the absence of a factor of  $2\pi$  in the numerator of the exponential, the Fourier transform of  $p_y(Y)$ . This expected value is clearly a function of  $t$  (which should *not* be interpreted as time—its use is merely convention, as is the definition (10.38) without  $2\pi$  where our own convention would put it), which we will denote  $\phi_y(t)$ , and which is called the “characteristic function”. Now suppose we have two *independent* random variables  $y, x$ . Consider the characteristic function for  $w = x + y$ ,

$$\phi_w(t) = \langle e^{i(x+y)t} \rangle = \langle e^{ixt} \rangle \langle e^{iyt} \rangle = \int_{-\infty}^{\infty} e^{iXt} p_x(X) dX \int_{-\infty}^{\infty} e^{iYt} p_y(Y) dY, \quad (10.39)$$

by independence. So

$$\phi_w(t) = \phi_x(t) \phi_y(t), \quad (10.40)$$

that is to say, the product of two Fourier transforms. Because characteristic functions are defined as Fourier transforms, it follows that the inverse transform is,

$$p_w(W) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_w(t) e^{-iWt} dt \quad (10.41)$$

with the  $1/2\pi$  being necessary because of its absence in the complex exponentials. But by the convolution theorem (or its very slight re-derivation with this changed convention), we must have

$$p_w(W) = \int_{-\infty}^{\infty} p_x(W') p_y(W - W') dW'. \quad (10.42)$$

The solution generalizes in an obvious way to the sum of an arbitrary number of independent variables. If  $p_x = p_y$ , we have  $\phi_w = \phi_y(t)^2$  or for the sum of  $n$  such variables,  $\phi_w(t) = \phi_y(t)^n$ .

The characteristic function of the  $G(0, 1)$  Gaussian is readily found to be

$$\phi(t) = e^{-t^2/2}, \quad (10.43)$$

and thus the sum of two  $G(0, 1)$  variables would have a characteristic function,  $e^{-t^2/2} e^{-t^2/2} = e^{-t^2}$ , whose inverse transform is found to be

$$p(X) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-X^2}, \quad (10.44)$$

that is to say another Gaussian of zero mean, but variance 2. It follows immediately from (10.43) that a sum of  $n$  -  $G(0, 1)$  variables is a new  $G(0, n)$  variable. If  $X$  is instead the square of a  $G(0, \sigma^2)$  variable, then Eq. (10.44) becomes

$$p(X) = \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma^2} e^{-X^2/\sigma^2} \quad (10.45)$$

One use of the characteristic function is in its relations to the moments of the corresponding probability density. Let us suppose  $\phi(t)$  has a convergent Taylor series about  $t = 0$  :

$$\phi(t) = 1 + t \left. \frac{d\phi(t)}{dt} \right|_{t=0} + \frac{t^2}{2!} \left. \frac{d^2\phi(t)}{dt^2} \right|_{t=0} + \dots \quad (10.46)$$

But from (10.38) we can evaluate the successive derivatives:

$$\begin{aligned} \left. \frac{d\phi(t)}{dt} \right|_{t=0} &= \int_{-\infty}^{\infty} iY p_y(Y) dY = im, \\ \left. \frac{d^2\phi(t)}{dt^2} \right|_{t=0} &= \int_{-\infty}^{\infty} (iY)^2 p_y(Y) dY = i^2 m_2, \\ \left. \frac{d^3\phi(t)}{dt^3} \right|_{t=0} &= \int_{-\infty}^{\infty} (iY)^3 p_y(Y) dY = i^3 m_3, \\ \left. \frac{d^{(n)}\phi(t)}{dt^{(n)}} \right|_{t=0} &= \int_{-\infty}^{\infty} (iY)^n p_y(Y) dY = i^n m_n, \end{aligned} \quad (10.47)$$

Where  $m_i$  are the successive moments of the probability density. Thus the successive moments determine the terms of the Taylor series expansion in (10.46), and hence the probability density itself. These results can be turned into a statement that a knowledge of all of the moments  $m_i$  of a random process is usually equivalent to knowledge of the complete probability density. Conversely, knowledge of the characteristic function means that all of the moments are readily generated from its derivatives evaluated at  $t = 0$ . (There are probability densities whose moments are not finite, e.g., the Cauchy, and the argument fails for them.)

The characteristic function generalizes to multivariate probability densities by introduction of Fourier transforms in several dimensions. Moment generation is achieved by using Taylor series in several dimensions.

### Central Limit Theorems (CLT)

Consider the sum of  $n$  independent variables,  $\xi_i$ ,

$$\xi = \xi_1 + \dots + \xi_n \quad (10.48)$$

all having the same mean,  $m_1$  and variance,  $\sigma_1^2$ , and with the same, but arbitrary, probability density,  $p_1(\Xi)$ . Then

$$\langle \xi \rangle = m = nm_1, \langle (\xi - m)^2 \rangle = \sigma^2 = n\sigma_1^2. \quad (10.49)$$

Define the normalized variable

$$\tilde{\xi} = \frac{\xi - m}{\sigma} = \tilde{\xi}_1 + \dots + \tilde{\xi}_n, \quad \tilde{\xi}_i = \frac{\xi_i - m_1}{\sqrt{n}\sigma_1}. \quad (10.50)$$

Suppose that the characteristic function of  $\xi_i$  is  $\phi_1(t)$ . Then by the shift and scaling theorems of Fourier transforms, the characteristic function of  $\tilde{\xi}_i$  is  $\tilde{\phi}_1(t) = \phi_1(t/(\sigma_1\sqrt{n})) \exp(-im_1t/(\sigma_1n))$ . Hence the characteristic function for  $\tilde{\xi}$  must be

$$\tilde{\phi}(t) = \tilde{\phi}_1(t)^n = \left[ e^{-im_1t/(\sigma_1\sqrt{n})} \phi_1(t/(\sigma_1\sqrt{n})) \right]^n. \quad (10.51)$$

Now  $\tilde{\phi}_1(t)$  is evidently the characteristic function of a random variable with zero mean and variance  $1/n$ . Thus it must be true that (expanding in a Taylor Series), and using (10.46, 10.47),

$$e^{-im_1t/(\sigma_1\sqrt{n})} \phi_1(t/(\sigma_1\sqrt{n})) = 1 - \frac{t^2}{2n} + O\left(\frac{t^3}{\sigma_1n^{3/2}}\right). \quad (10.52)$$

Thus to lowest order, (10.51) is

$$\tilde{\phi}(t) = \left(1 - \frac{t^2}{2n}\right)^n. \quad (10.53)$$

Taking the limit as  $n$  goes to infinity, and invoking L'Hôpital's rule on the log of  $\tilde{\phi}(t)$ , we have

$$\tilde{\phi}(t) \rightarrow e^{-t^2/2} \quad (10.54)$$

Thus in this limit, the probability density for  $\tilde{\xi}$  is the inverse transform of (10.54) and is

$$\tilde{p}(\Xi) = \frac{1}{\sqrt{2\pi}} e^{-\Xi^2/2} \quad (10.55)$$

that is,  $G(0, 1)$ . Or, using the scaling and shift theorems,

$$p(\Xi) = \frac{1}{\sqrt{n}\sigma_1\sqrt{2\pi}} e^{-(\Xi - nm_1)^2/(2\sqrt{n}\sigma_1)}. \quad (10.56)$$

That is to say, we have shown that sums of large numbers of random variates have a tendency to become Gaussian, whether or not the underlying probability densities are themselves Gaussian. Result (10.56) is a special case of the so-called Central Limit Theorem (CLT), which can be proved under much more general circumstances. There are clearly restrictions that could prevent the limit (10.54) from being reached, but the CLT is often valid. Note the special case of a sample average,

$$\tilde{m} = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad (10.57)$$

for which it follows immediately that  $\tilde{m}$  will have a probability density  $G(m, \sigma^2/n)$ .

*Exercise.* Suppose that  $p_1(\Xi) = 1/2 \{\delta(\Xi - 1) + \delta(\Xi + 1)\}$ . Study the behavior of (10.51) as  $n \rightarrow \infty$ . What is the limiting probability density of the sum? Hint: use the binomial expansion on  $(1/2^n) (e^{it} + e^{-it})^n$ .

### Stationarity

Consider a  $G(0, \sigma^2)$  time series  $x_m$  such that  $R_{nm} = \langle x_n x_m \rangle$ . Given the zero mean, and specified  $\mathbf{R}$ , we infer that the joint probability density  $p_{x_1 x_2 \dots x_k}$  is (10.23) with  $\mathbf{m} = \mathbf{0}$ . Let us suppose that  $R_{nm}$  depends only upon the time difference  $\tau = n - m$ , so that  $R_{nm} = \langle x_n x_{n+\tau} \rangle = R_\tau$ , independent of  $n$ . Such a time series is said to be "stationary in the wide sense". If it is also true that all statistics, including all higher moments such as  $\langle x_n x_m x_p \rangle$ , etc. only depend upon the time interval among the

time series elements, the time series is said to be “stationary” or “stationary in the strict sense”. It is another nice property of normal variates that if they are stationary in the wide-sense, they are stationary in the strict sense (most readily seen by observing that in (10.23)  $\mathbf{R}$  has all of its parameters dependent solely upon  $\tau$ , and not upon the absolute time. Hence, the probability density depends only upon  $\tau$ , as does any statistical property derived from it. The theory for stationary time series is highly developed, and it is commonly assumed that one is dealing in nature with stationary processes, but one must be alert to the potential failure of the assumption.

### Sample Estimates

In working with random variables, it is very important to distinguish between a theoretical value, e.g., the true average, written  $\langle y \rangle$ , or  $m$ , and the *sample* value, such as the sample average, written as either  $\langle y \rangle_N$  or  $\tilde{m}$ , where the  $N$  subscript indicates that it was based upon  $N$  observations. We use a tilde,  $\sim$ , to mark an estimate of a variable; estimates are themselves always random variables (e.g.  $\tilde{m}$ ) where the parameter itself,  $m$  is not.

The usual sample average is

$$\tilde{m} = \langle y \rangle_N = \frac{1}{N} \sum_{i=1}^N y_i. \quad (10.58)$$

A useful property of a sample mean (or *estimator* of the mean) is that it's own expected value should be equal to the true value (this is not always true, and as will be seen, it is not always completely desirable.) For example,

$$\langle \tilde{m} \rangle = \frac{1}{N} \sum_{i=1}^N \langle y_i \rangle = \frac{1}{N} \sum_{i=1}^N m = m \quad (10.59)$$

Such an estimator is said to be unbiased. (If we could average  $N$  experimental values in  $M$  separate experiments, the average of the sample averages would be expected to be the true average.)

It helps to know what is the scatter of an estimate about its true value, so consider the variance:

$$\langle (\tilde{m} - m)^2 \rangle = \left\langle \left[ \frac{1}{N} \sum_{i=1}^N y_i - m \right]^2 \right\rangle = \left\langle \left[ \frac{1}{N} \sum_{i=1}^N (y_i - m) \right]^2 \right\rangle \quad (10.60)$$

$$= \frac{1}{N^2} \sum_{i=1}^N \langle (y_i - m)^2 \rangle = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N} \quad (10.61)$$

so that the standard deviation of a sample mean about the true mean is  $\sigma/\sqrt{N}$ , which is the famous “square-root of  $N$ ” rule.

Now consider the sample variance:

$$\tilde{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \tilde{m})^2 \quad (10.62)$$

computed as the sample variance about the sample mean. A little algebra (left as an exercise), shows that,

$$\langle \tilde{\sigma}^2 \rangle = \frac{N-1}{N} \sigma^2 \neq \sigma^2 \quad (10.63)$$

that is to say, the sample variance is *biased* (although it is asymptotically unbiased as  $N \rightarrow \infty$ ). The bias is readily removed by re-defining

$$\tilde{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \tilde{m})^2. \quad (10.64)$$

The origin of the bias in this case is that in (10.62) the  $N$  terms being averaged are not independent of each other; rather since the sum is taken over the values of  $y_i - \tilde{m}$ , the last term could be predicted from the preceding  $N - 1$  of them by the requirement that the sample average actually is  $\tilde{m}$ . Consequently there are only  $N - 1$  independent terms ( $N - 1$  “degrees-of-freedom”) in the sum.

### Consistent Estimators

Data are used to estimate various properties characterizing the statistical population from which the observation come. For example, the sample mean (10.58), is intended to be a good estimate of the true average  $m$ , from which the data were drawn. Many other properties are estimated, including the variance and the power density spectrum (defined below). Any estimate of a property has to be examined for at least two desirable properties: (1) that the average value of the estimate should be the true value (for an unbiased estimate), at least as the number of samples becomes large, and (2) that as the number of samples becomes large, the variance of the sample about its mean value (which one might hope is the true value) ought to go to zero. Estimators with these two properties are said to be “consistent”. Sometimes however, demanding an unbiased estimator greatly increases the variance of the estimate, and one may deliberately permit a bias if the variance can thereby be reduced. The sample mean was shown above to be unbiased, and by (10.60) its variance about the true value diminishes with  $N$ ; hence it is a consistent estimator. One can also show that  $\tilde{\sigma}^2$  in the form (10.64) is also a consistent estimator as is the original definition, in the asymptotic limit.

A useful idea in this context is the Chebyshev inequality. Let  $\xi$  be any random variable with true mean  $m$ , variance  $\sigma^2$  and second moment  $\langle \xi^2 \rangle = m_2$  (that is not taken about the mean). Then

$$m_2 = \int_{-\infty}^{\infty} \Xi^2 p(\Xi) d\Xi \geq \left( \int_{-\infty}^{-\delta} + \int_{\delta}^{\infty} \right) \Xi^2 p(\Xi) d\Xi \quad (10.65)$$

$$\geq \delta^2 \left( \int_{-\infty}^{-\delta} + \int_{\delta}^{\infty} \right) p(\Xi) d\Xi \quad (10.66)$$

since  $\xi^2$  is always greater than or equal to  $\delta^2$  in the integral. So we have the weak inequality

$$\int_{|\xi| > \delta} p(\Xi) d\Xi \leq \frac{m_2}{\delta^2}, \quad (10.67)$$

which we can read as “the probability that  $|\xi| > \delta$  is less than or equal to  $m_2/\delta^2$ .” If we replace  $\xi$  by  $\xi - m$ , and then  $m_2 = \sigma^2$ , we have

$$\text{prob} \{ |\xi - m| \geq \delta \} \leq \frac{\sigma^2}{\delta^2} \quad (10.68)$$

where “prob” denotes “the probability that”. The sense of the inequality can be inverted to

$$\text{prob}\{|\xi - m| \leq \delta\} \geq 1 - \frac{\sigma^2}{\delta^2}, \quad (10.69)$$

which is the Chebyshev inequality. These inequalities in turn lead to the important idea of “convergence in probability”.  $\xi_k$  (where  $k$  is a parameter) is said to converge in probability to  $c$ , as  $k \rightarrow \infty$ , if  $\text{prob}\{|\xi_k - c| \geq \delta\} \rightarrow 0$ . This is convergence that is “almost surely”, written a.s., and differs from ordinary mathematical convergence in being a statement that deviation from the asymptotic value becomes extremely improbable, but not impossible, in the limit.

Let us apply the Chebyshev inequality to the sample mean (10.58), whose variance is  $\sigma^2/N$ . Then by (10.69),

$$\text{prob}\{|\tilde{m} - m| \geq \delta\} \leq \frac{\sigma^2}{N\delta^2} \quad (10.70)$$

Thus as  $N \rightarrow \infty$  (corresponding to the parameter  $k$  above), the probability that the sample mean differs from the true mean by any amount  $\delta$  can be made arbitrarily small. It is thus a consistent estimator. For the spectral estimators, etc. used below, one needs formally to show convergence in probability to the true values, but this demonstration is normally left to the reader.

#### Confidence Intervals

Consider the sample mean  $\tilde{m}$ , whose own mean is  $m$  and whose variance is  $\sigma^2/N$ . If it is derived from  $x_i$  which are normal independent, identically distributed (i.i.d) variates, then  $\tilde{m}$  is  $G(m, \sigma^2/N)$ . Let us make a table of the canonical  $G(0, 1)$  variable:

$\eta(\alpha/2)$	$\text{prob}[-\eta \leq X \leq \eta]$	$\alpha$
1.0	0.683	0.317
1.96	0.950	0.050
2.00	0.954	0.046
2.58	0.990	0.010
3.00	0.997	0.003

Table Caption. (Taken from Jenkins and Watts, 1968, p. 71). Here  $\eta$  is the value, symmetric about the mean of 0, between which the probability is  $1 - \alpha$  that a random sample  $X$  would lie.  $\alpha$  is the fraction of the value which would lie outside the range  $\pm\eta$ . Thus a random sample  $X$  would have a probability of  $0.95 = 1 - \alpha$  of lying in the range  $\pm 1.96$ . In many trials, one would expect 5% of the values to lie, by chance, outside this range. The normal density is symmetric about the mean, and which is not true for more general densities.

---

Then

$$\text{prob}\left\{-\eta(\alpha/2) \leq \frac{\tilde{m} - m}{\sigma/\sqrt{N}} \leq \eta(\alpha/2)\right\} = 1 - \alpha \quad (10.71)$$

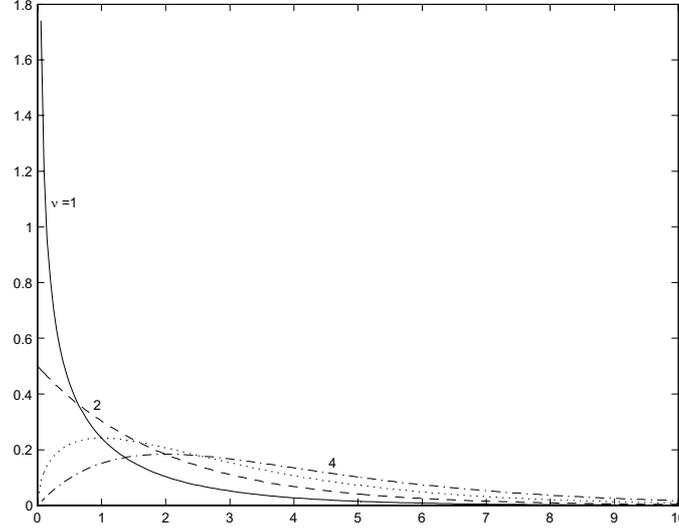


FIGURE 9. Probability density for  $\chi_\nu^2$ ,  $\nu = 1, \dots, 4$ . Note growth of long positive tail as  $\nu$  increases.  $m = \nu$ , and  $\sigma^2 = 2\nu$ . For large  $\nu$ ,  $\chi_\nu^2$  approaches a normal probability density. The cases  $\nu = 1, 2$  are exceptional in having their maximum at 0.

can be read from the table for any given  $\alpha$ . We can re-arrange (10.71) to:

$$\text{prob} \left\{ m - \frac{\eta(\alpha/2)\sigma}{\sqrt{N}} \leq \tilde{m} \leq m + \frac{\eta(\alpha/2)\sigma}{\sqrt{N}} \right\} = 1 - \alpha \quad (10.72)$$

Alternatively,

$$\text{prob} \left\{ \tilde{m} - \left( \sigma/\sqrt{N} \right) \eta(\alpha/2) \leq m \leq \tilde{m} + \left( \sigma/\sqrt{N} \right) \eta(\alpha/2) \right\} = 1 - \alpha. \quad (10.73)$$

This last form is the most useful one: suppose to be specific that  $\alpha = .05$ . Then for a sample mean  $\tilde{m}$  obtained from a random variate having variance  $\sigma^2$ , the probability is 0.95 that the true value,  $m$ , lies in the interval  $\tilde{m} \pm \sigma \eta(.05/2)$ . This interval is said to be a  $1 - \alpha\%$  confidence interval (here a 95% confidence interval). It clearly diminishes to zero as  $N \rightarrow \infty$ .

Consider now a non-symmetric probability density. One that proves very useful to us is the so-called  $\chi_\nu^2$  (chi-square with  $\nu$  degrees-of-freedom),

$$\begin{aligned} p_\xi(X) &= \frac{1}{2^{\nu/2} \Gamma(\nu/2)} X^{\nu/2-1} \exp(-X/2), \quad X > 0 \\ &= 0, \quad X \leq 0 \end{aligned} \quad (10.74)$$

whose mean is  $\nu$ , and variance is  $2\nu$ . (This probability density describes a variable  $x = \sum_1^\nu \xi_i^2$ , where the  $\xi_i$  are independent,  $G(0, 1)$ .) It is plotted in fig. 9 for  $1 \leq \nu \leq 4$ . For present purposes, we note only that the probability density is non-symmetric about its mean, having a long tail toward high positive values. Consider the tail at the low end containing a fraction  $\alpha/2$  of the values from a set of random

trials, and denote the value  $X$  below which this tail falls as  $\eta_-(\alpha/2)$ . Correspondingly, the tail at the high end occupying  $\alpha/2$  of the trial results is to lie to the right of  $X = \eta_+(\alpha/2)$ . Suppose now that we have a variable  $\xi$  which has been estimated as  $\tilde{\xi}$  and which is thought to be distributed in  $\chi_\nu^2$  with mean  $\nu$  and variance  $2\nu$ . Then

$$\text{prob} \left\{ \eta_-(\alpha/2) \leq \tilde{\xi} \leq \eta_+(\alpha/2) \right\} = 1 - \alpha. \quad (10.75)$$

To employ this in a practical example, consider the sample variance  $\tilde{\sigma}^2$  (10.64) constructed from  $N$  identically distributed variables, which are  $G(m, \sigma^2)$ . It is easy to show (sketched out below) that  $(N-1)\tilde{\sigma}^2/\sigma^2$  will be a  $\chi_\nu^2$  variable with  $\nu = N-1$  (it's expected value is  $N-1$ , and its variance is  $2(N-1)$ ). Then setting  $\tilde{\xi} = (N-1)\tilde{\sigma}^2/\sigma^2$  in (10.75) and rearranging, we obtain

$$\text{prob} \left\{ \frac{\tilde{\sigma}^2(N-1)}{\eta_+(\alpha/2)} \leq \sigma^2 \leq \frac{\tilde{\sigma}^2(N-1)}{\eta_-(\alpha/2)} \right\} = 1 - \alpha \quad (10.76)$$

and the true variance would lie between these two (unequal) bounds about  $\tilde{\sigma}^2$ . Because  $\eta_- < \eta_+$ , the upper limit of the  $\alpha\%$  confidence limit will be further above  $\sigma^2$  than the lower limit will be below (for  $\nu > 2$ ).  $\eta_\pm$  are tabulated in various statistics books or can be calculated from various software packages. (It is sometimes useful to have the  $\chi_\nu^2$  distribution for the sum,  $\xi$ , of squared zero-mean Gaussian variates of variance  $\sigma^2$ , or:

$$\begin{aligned} p_\xi(X) &= \frac{1}{\sigma^2 2^{\nu/2} \Gamma(\nu/2)} \left( \frac{X}{\sigma^2} \right)^{\nu/2-1} \exp(-X/(2\sigma^2)), \quad X \geq 0 \\ &= 0, \quad X < 0. \end{aligned} \quad (10.77)$$

In examining the form (10.73), one might object that  $\sigma^2$  is not likely to be known. This led Gosset (writing under the famous pseudonym of "Student") to show that he could find the probability density of the variable  $T_{N-1} = \sqrt{N}(\tilde{m} - m)/\tilde{\sigma}$  and which is not dependent upon  $\sigma$ . The resulting probability density is called Student's  $t$ -distribution. We leave its discussion to the references (see, e.g., Cramér, 1946, Section 18.2).

### White Noise

A white noise process is a stationary time series (sequence) with a uniform in time variance, zero mean, and in which knowledge of the value at one time  $\theta_m$  carries no information about its value at any other time, including the immediately preceding and following times. That is,  $\langle \theta_m \theta_{m'} \rangle = \sigma_\theta^2 \delta_{mm'}$ . A sequence of coin flips is such a process. The general terminology is that these are independent identically distributed variables (or an i.i.d). Often, we will assume that the values are normally distributed, as in (10.2) with  $m = 0$ .

White noise is the simplest possible stochastic process. Let us therefore consider its Fourier transform or series using the real form (6.29-6.31). The coefficients are

$$a_k = \frac{2}{N} \sum_{p=0}^{N-1} \theta_p \cos\left(\frac{2\pi kp\Delta t}{T}\right), k = 0, \dots, N/2, \quad (10.78)$$

$$b_k = \frac{2}{N} \sum_{p=0}^{N-1} \theta_p \sin\left(\frac{2\pi kp\Delta t}{T}\right), k = 1, \dots, N/2 - 1. \quad (10.79)$$

It follows immediately that

$$\begin{aligned} \langle a_k \rangle &= \langle b_k \rangle = 0 \\ \langle a_k^2 \rangle &= \frac{4}{N^2} \left\langle \sum_{p=0}^{N-1} \theta_p \cos\left(\frac{2\pi kp\Delta t}{T}\right) \sum_{r=0}^{N-1} \theta_r \cos\left(\frac{2\pi kr\Delta t}{T}\right) \right\rangle \\ &= \frac{4}{N^2} \sum_p \sum_r \langle \theta_p \theta_r \rangle \cos\left(\frac{2\pi kp\Delta t}{T}\right) \cos\left(\frac{2\pi kr\Delta t}{T}\right) \\ &= \frac{4}{N^2} \sum_p \sum_r \delta_{pr} \sigma_\theta^2 \cos\left(\frac{2\pi kp\Delta t}{T}\right) \cos\left(\frac{2\pi kr\Delta t}{T}\right) \\ &= \frac{2}{N} \sigma_\theta^2 \end{aligned} \quad (10.80)$$

by (6.30). Similarly,

$$\langle b_k^2 \rangle = \frac{2}{N} \sigma_\theta^2 \quad (10.81)$$

$$\langle a_n b_k \rangle = 0 \quad (10.82)$$

$$\langle a_k a_n \rangle = \langle b_k b_n \rangle = \frac{2}{N} \delta_{kn} \sigma_\theta^2, \quad (10.83)$$

omitting the zero and Nyquist frequencies. For these frequencies,

$$\langle a_0^2 \rangle = \langle a_{N/2}^2 \rangle = \frac{4}{N}. \quad (10.84)$$

To say this in words: the Fourier transform of a white noise process has zero mean and is uncorrelated from one frequency to another; the sine and cosine amplitudes are uncorrelated with each other at all frequencies, and the variance of the sine and cosine components is uniform with frequency. If the  $\theta_m$  are normally distributed,  $G(0, \sigma_\theta^2)$  then it follows immediately that  $a_k, b_k$  are also normally distributed  $G(0, 2\sigma_\theta^2/N)$ . The Parseval relationship requires

$$\frac{1}{N} \sum_{m=0}^{N-1} \theta_m^2 = \frac{a_0^2}{4} + \frac{1}{2} \sum_{m=1}^{N/2-1} (a_m^2 + b_m^2) + \frac{a_{N/2}^2}{4}, \quad (10.85)$$

here including the mean and Nyquist frequencies. (The true mean is zero, but the actual sample mean will not be, although it is often set to zero for numerical reasons. Doing so, means that only  $N - 1$  of

the terms on the left in (10.85) would then be independent.) To check this last equation, we can see if it holds on the average:

$$\begin{aligned} \frac{1}{N} \sum_{m=0}^{N-1} \langle \theta_m^2 \rangle &\stackrel{?}{=} \frac{\langle a_0^2 \rangle}{4} + \frac{1}{2} \sum_{n=1}^{N/2-1} \langle (a_n^2 + b_n^2) \rangle + \frac{\langle a_{N/2}^2 \rangle}{4}, \text{ or,} \\ \sigma_\theta^2 &\stackrel{?}{=} \frac{\sigma_\theta^2}{N} + \frac{1}{2} \sum_{n=1}^{N/2-1} \left( \frac{2}{N} + \frac{2}{N} \right) \sigma_\theta^2 + \frac{\sigma_\theta^2}{N} = \frac{\sigma_\theta^2}{N} + \frac{1}{2} \left( \frac{N}{2} - 1 \right) \frac{4}{N} \sigma_\theta^2 + \frac{\sigma_\theta^2}{N} = \sigma_\theta^2, \end{aligned} \quad (10.86)$$

as required.

As the record length grows, the number  $N$ , of Fourier coefficients grows linearly, but the mean square power  $(1/N) \sum_{m=0}^{N-1} \theta_m^2$  in (10.85) remains fixed, independent of  $N$ . Thus the expected value of any  $a_n^2 + b_n^2$  is reduced by the factor  $1/N$  to compensate for their growing population.

If one computes the phase of the Fourier series as

$$\phi_n = \tan^{-1}(b_n/a_n) \quad (10.87)$$

it is readily seen that  $\phi_n$  has a uniform probability density

$$p_\phi(\Phi) = \frac{1}{2\pi}. \quad (10.88)$$

(To see this algebraically, we recognize that if  $\theta_m$  are  $G(0, \sigma_\theta^2)$ ,  $a_n, b_n$  are uncorrelated Gaussian variables distributed  $G(0, \sigma_\theta^2/2N)$  and hence they are also independent. Because they are independent, their joint probability density is the product of their identical probability densities:

$$\begin{aligned} p_{a_n, b_n}(\Xi_1, \Xi_2) &= p_{a_n}(\Xi_1) p_{b_n}(\Xi_2) = \\ &= \frac{1}{2\pi(\sigma_\theta^2/2N)} \exp(-\Xi_1^2/(\sigma_\theta^2/2N)) \exp(-\Xi_2^2/(\sigma_\theta^2/2N)) \end{aligned} \quad (10.89)$$

We can thus conclude that there is no information content in the phase of a white noise process. What then is the probability density for  $a_n^2, b_n^2$  and  $a_n^2 + b_n^2$ ? The probability densities for the first two must be identical, and are thus the densities for the square of a normal variate with 0 mean and variance  $\sigma_\theta^2/2N$ . Let us use normalized variables so that they have unit variance, i.e., consider  $a'_n = a_n/(\sigma_\theta/\sqrt{2N}), b'_n = b_n/(\sigma_\theta/\sqrt{2N})$  each of which will be  $G(0, 1)$ . Using the rule for change of variable we find that,

$$p_{a_n'^2}(X) = p_{b_n'^2}(X) = \frac{X^{-1/2}}{\sqrt{2\pi}} \exp(-X/2) \quad (10.90)$$

which is again  $\chi_1^2$  (recall Eq. (10.15)). Invoking the machinery for sums of independent variables, we find that the probability density for  $r_n'^2 = a_n'^2 + b_n'^2$  is

$$p_{r_n'^2}(X) = \frac{1}{2} \exp(-X/2) \quad (10.91)$$

which is called  $\chi_2^2$  (chi-square with two-degrees-of-freedom) and whose mean is 2 and variance is 4 (one has to do the integral; see Eq. 10.74).

## 1. FREQUENCY DOMAIN FORMULATION

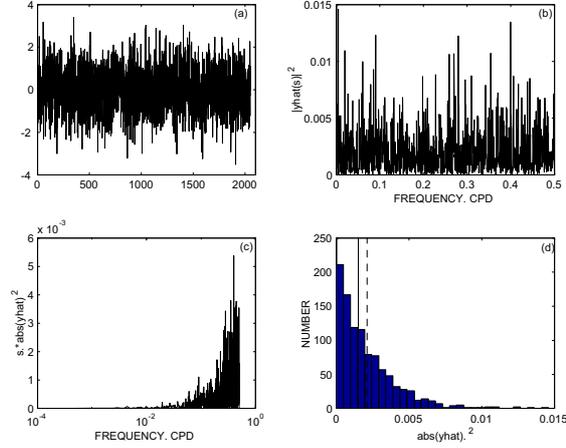


FIGURE 10. (a) 2048 numbers from a pseudo-random generator. These approximate a white noise process. (b) Linear scale periodogram of the values in (a). (c) Semi-logarithmic scale plot of  $s|\alpha_n|^2$  in which the multiplication by  $s$  compensates for the logarithmic squeezing of the frequencies at the high end. Notice the visual impression that energy increases with increasing frequency—an artifact of the plotting method. (d) 30 bin histogram of the values  $|\alpha_n|^2$  which should and do approximate a  $\chi_2^2$  variate. The mean (dashed) and median (solid) values are shown. There is evidently a much greater chance of a value far larger than the mean to occur than there is for a much smaller one to occur, owing to the limiting value at zero. But more smaller values below the mean occur than above the mean of  $\chi_2^2$ .

*Exercise.* Using the results of the exercise on P. 37, find the probability density for the amplitude and phase of the Fourier coefficients.

Fig. 10 shows the Fourier coefficients of a pseudo-random  $\theta_n$ , their sum of squares,  $r_n^2 = a_n^2 + b_n^2$ , and the histogram of occurrences of  $r_n^2$  along with the theoretical  $\chi_2^2$  (scaling the variables by  $\sigma_\theta/2N$  to restore the original variance). The true mean would be

$$\langle r_n^2 \rangle = 2 \frac{\sigma_\theta^2}{2N} = \frac{\sigma_\theta^2}{N} \quad (10.92)$$

and the variance of  $r_n^2$  would be

$$\langle (r_n^2 - \langle r_n^2 \rangle)^2 \rangle = 2 \left( \frac{\sigma_\theta^2}{N} \right)^2 = \left( \frac{\sigma_\theta^2}{N} \right)^2. \quad (10.93)$$

That is, the variance is proportional to the square of the of the power density.

*Definition.*  $r_n^2 = a_n^2 + b_n^2 = |\alpha_n|^2$  is called the “periodogram” of a random process.

REMARK 1. *There is a common, naive, assertion about the results of Fourier transforming a stochastic process. Basically, it says that “obviously, the low frequency Fourier components are less-well determined than are the high frequency ones, because many more oscillations of the high frequency components exist in the record, whereas for the first record harmonic,  $s_1 = 1/T$ , only one cycle is covered by the data.” Unfortunately this seemingly obvious conclusion is false. Consider that every Fourier harmonic,  $s_n$ , differs from its two neighboring harmonics,  $s_{n-1}, s_{n+1}$  by exactly one cycle per record length. All of the data are required to determine the Fourier amplitudes of these neighbors, separating them by consideration of their difference by one cycle in the record.  $s_1$  differs exactly in the same way from  $s_0$  (the record mean), and  $s_2$ . Thus precisely the same amount of information is available about  $s_1$  as about any other  $s_n$ . This statement is consistent with (10.93): the variance of the periodogram values is independent of  $n$ .*

It appears from (10.93) that the variance of the estimates diminishes with  $N$ . But the ratio of the variance to the estimate itself, is independent of  $N$ . That is, because the number of Fourier coefficients increases with  $N$ , each has a diminishing proportional contribution to the constant record variance (power), and the variability of the estimate as a fraction of the true expected value remains unchanged even as the record length goes to infinity. This behavior confused scientists for years.