Chapter 3

Probability

3.1 Why Probabilistic Descriptions?

If probability is only an approximate description of something more complicated, why not solve the problem correctly and forget probability? One could answer quite simply that probabilistic descriptions are necessary because of noise, but this statement in and of itself raises the additional question as to what noise is. This question is more than philosophical, as we will have to know a considerable amount about the causes of the noise we observe in order to be able to analyze it and eventually find ways to overcome it.

For purposes of this course, we will generally think of noise as being either thermal or quantum mechanical in nature. These two forms are not always separable from each other, although for the present we will try to separate them.

Let us briefly consider the nature of quantum noise before entering into a more broad-ranging discussion of thermal noise. These topics will also be the main themes of later chapters, namely Chapters 7 (quantum) and 8 (thermal). Quantum noise is usually thought of as arising when one wants to know a very precise answer to some question—for example, when will an electron strike a ground plane. (A more complicated question might be, what is an AC ground plane?, but more on that later.) If one knew the exact state of the motion of an electron under the action of some force at some time, one could calculate the arrival time. That is, if one knew the force $F$ an electron would experience, one could write (approximately) that

$$F = m_e a_e = m_e \ddot{x}_e,$$

where $m$ is the mass and $x$ is the vector position. To solve such an equation as (3.1), however, would require that one be able to specify the initial position $x_e(0)$ and velocity $\dot{x}_e(0)$ of the electron. Quantum mechanics, however, tells us we cannot really do this, as the electron has extents in both physical space and momentum (transform) space such that its wave function’s standard deviations must satisfy

$$\Delta x \Delta p \geq \frac{\hbar}{4\pi} = \frac{\hbar}{2},$$

where $\hbar = \hbar / 2\pi$ is Planck’s reduced constant. When the equality holds, we say that we have a coherent state. Equation (3.2) states that we cannot simultaneously know both an electron’s position and its momentum, but a further quantum mechanical consequence is that trying to measure one or the other with any precision changes the state of the electron with the consequence that both variables are altered. Measurements designed to specify one variable precisely must by nature alter the conjugate variable’s value severely. Trying to precisely localize an electron by hitting it with a second high-velocity electron will give a value of position. However, it will not only erase the original velocity but impart a new and large unknown one. Multiple measurements cannot, therefore, be used to reduce the uncertainty bars.

One could probably argue that we do not really need to specify the time that each electron arrives at a given place. Maybe we only need to know on average when a large number arrive. But we are discussing noise. The point is to know how close we can come to predicting something exactly—at what point our measurement apparatus becomes so good that we see junk on our signal. The $\Delta x \Delta p$ argument just presented has the
consequence that at no point can we specify the initial state of a system with infinite accuracy. The question arises, therefore, as to whether there are systems which are sufficiently poorly behaved that we would need such precision. This question we will take up in a later paragraph purportedly addressing thermal noise. At present, let us look at another aspect of uncertainty.

Another manner of writing the uncertainty relation of (3.2) is in an energy representation:

\[ \Delta E \Delta t \geq \frac{\hbar}{2}, \]  

(3.3)

where equality holds for a coherent state. The consequence that we cannot really know the time of an event too precisely is not too surprising in light of the above discussion of position and momentum. Again, the question arises as to when this uncertainty could be important. It turns out that, for atomic energy level transitions whose energy corresponds to optical photons, the time uncertainties are quite important and, in fact, at typical power levels correspond to typical rates in optical communications systems. Such noise is generally referred to as shot noise. Perhaps an equally important consequence is that a coupled quantum mechanical system can spontaneously exchange energy between its subsystems; i.e., the atomic state can be lowered through emission of a photon into the optical subsystem. This spontaneous emission process is an important one in that it is the one that makes the sun and stars shine as well as serving as the basic principle of the electric light bulb (although Edison discovered it without the help (or hindrance) of quantum electrodynamics).

Clearly, the above-discussed quantum noise sources of shot noise and spontaneous emission noise are fundamentally random due to their quantum nature. Does all that we refer to as random noise arise from these fundamentally random causes, or is there noise that we call random that really is describable deterministically but whose description is too complex for us to carry out?

The primary source of thermal noise as we know it is that if we bang something it shakes. In a complicated mechanical system, enough shaking will lead to seemingly random internal motion. How random is this internal motion? Let us consider a closed system with some set of initial conditions. Let us further assume that the system can be described by Newton’s laws. Now if we truly assume the system is closed and the containment walls are lossless, the equations coupling the motions of the \( N \) particles in the system should take the form

\[ \ddot{x}_i = \sum_{j=1}^{N} M_{ij} x_j, \]  

(3.4)

where the matrix \( M \) represents all of the particle interactions which we have assumed here are only position-dependent. A troubling point about this set of equations is that, if we make the transformation \( t \rightarrow -t \) in them, we get back the same set of equations. Why should this be troubling?

Let us say that we were to define an outrageous initial condition for our set of equations of (3.4) such that at time \( t = 0 \) all of the particles were hunched together in a little corner of our container. If we let them loose, then they would fill the container. However, the reversibility argument states that, if we started the system in one of the states where the particles filled the container, the system would be equally likely to collapse back to a corner or stay filling the container, all depending on an arbitrary sign of the initial condition.

The answer to the reversibility problem really lies within the realm of statistical mechanics. (See, for example, Kittel and Kroemer 1980.) The entropy of a system is really a measure of the number of states available to a system as a function of the total energy of a system. Clearly, a more confined system (e.g. less total volume) will have fewer configurations available to achieve a given energy. The particles hunched in a corner will have many less states available than a system spread throughout a volume. When one takes into account typical material densities (10\(^{23}\)/cm\(^3\)) versus typical macroscopic sizes (centimeters or greater) and also the fact that numbers of states tend to be factorials of numbers of particles, one can rapidly see why the second law of thermodynamics holds. The point is that, at standard temperatures, particles move at given velocities. The system will wander through states of roughly constant energy, of which there are very many, in phase space (a space with position and velocity axes with an axis for each of the three positional coordinates of each of the particles) at some rate. If we assume that the wandering is approximately random and that the probability of being at a location in phase space is proportional to the number of states there,
one will find that the average time it takes to enter a “low” probability state is so long that it cannot occur during the lifetime of our universe ($10^{15}$ years). An interesting example is given in Kittel and Kroemer 1980, p. 53. Therefore, reversibility versus irreversibility is really a question of how long one has to wait. The second law of thermodynamics really contains an implicit time constant.

But what does all of this say about noise? If we had a single particle in a box at temperature $T$, we would have a case where the particle could exist in any of a discrete set of states—that is, at different discrete points in the phase space of position and momentum, which are discrete due to the boundary conditions at the walls and the quantum nature of the particle, causing the momentum and position to satisfy an uncertainty principle. If these states were separated by an energy difference less than $kT$ ($k$ being Boltzmann’s constant), the particle could make a transition between them. That is, the particle could borrow (or give back) a quantum of heat to change its position/momentum state. If one were to put another identical particle in the box, there would be more states (probably double) more closely spaced. If one puts $N$ particles in the box, there will be $N!$ times the number of states spaced $1/N!$ times as far apart in energy space. With $N$ being on the order of $10^{20}$ or more, these states will be truly closely spaced. No matter how small $kT$ may be, it will cause the system to drift through a multiplicity of states continuously in time. But this really is what we mean by noise. The system is constantly changing its configuration. It will tend to stay in the “highest probability” (“highest entropy”) states, but there are so many of these that by nature this does not limit the uncertainty but actually maximizes it. Could one really predict the behavior of this system were one to have the “exact” initial conditions? The answer again lies in the enormity of states. There are so many states available that there is not enough computer power or time available in this universe to make the computation. In optical communications, we are dealing with a tiny microscopic corner of a big macroscopic box whose behavior it is beyond our ability to predict. Hence, our situation is noisy.

A last yet interesting point concerning the probability description issue of this section relates back to the earlier discussion of initial conditions. Do practical systems have outputs that stay bounded within uncertainty limits, limits defined by the uncertainty of the inputs? The answer is a resounding “no” as it turns out. Clearly, in any real system there should be nonlinear terms included in the system of Equation (3.4). Studies of the effects of such terms really go back to ones carried out by Fermi, Pasta, and Ulam at Los Alamos right after the war. (See, for example, the review article, May 1976.) Their result was that nonlinear terms did not lead to equipartition of energy but led to strange periodic behavior that was later attributable to solitons. That case was a special one, though. The more general nonlinear case really relates to some later work done at Los Alamos by Feigenbaum. He was able to show that even very simple nonlinear systems can exhibit very complex dynamical behavior. This behavior includes a limit known as chaos, wherein the system solutions diverge exponentially from each other for even infinitesimally close initial conditions. Such behavior in nonlinearly coupled systems now seems to be ubiquitous as more and more cases are uncovered.

The conclusions of the somewhat drawn-out overview discussion are that probabilistic descriptions really are necessary. Both quantum and thermal noise can give rise to truly random time evolution—quantum noise through shot noise and spontaneous emission and thermal noise through complexity, complexity which erases initial conditions which may be random already through quantum effects, both through entropy as well as nonlinear dynamical effects.

The following subsections will present the rudiments of the language of probability, following this long-winded justification. Much of the material in the following is also covered in a more terse manner in the appendices of Gagliardi and Karp (1988).

### 3.2 Simple Random Variables

#### 3.2.1 Probability Density

Let $x$ be a random variable. Then $x$ is a number out of a set of numbers, and there is a certain probability that $x$ is in an interval $I$. This probability is given by

$$\text{Prob}(x \in I) = \int_I p_x(x) \, dx,$$  \hspace{1cm} (3.5)
where $p_x(x)$ is the probability density function of $x$. This means that, if we measure $x$ $N$ times and let $N$ be a large number such that $N \to \infty$, we can plot the density function of $x$ as illustrated in Figure 3.1. The $N$ measured values of $x$ are called an ensemble.

The distribution function of $x$ is the probability that $x$ is smaller than some given value:

$$P(x) = \text{Prob}(x \leq x_i).$$  \hfill (3.6)

The probability density and the distribution functions are related by

$$p_x(x) = \frac{dP(x)}{dx}. \hfill (3.7)$$

If the distribution function $P(x)$ is continuous, we say that the random variable $x$ is continuous. In this case, the probability of $x$ being exactly some $x$ is zero:

$$\text{Prob}(x = x_i) = 0 \text{ for every } x. \hfill (3.8)$$

If $P(x)$ is a staircase function, then $x$ is a discrete random variable and $\text{Prob}(x = x_i) = p_i$, as in Figure 3.2. A random variable can also be mixed. For example, let $x$ be a 0 or a 1 (digital):

$$x_I(t) = \begin{cases} 0, & t \notin I \\ 1, & t \in I, \end{cases} \hfill (3.9)$$

then $\text{Prob}(x = 1) = p(1)$ and $\text{Prob}(x = 0) = 1 - p(1)$, and this is a discrete event although the space of $I$ can be infinite.

To summarize, the main properties of the distribution and probability density functions are that

1. $P(x)$ is monotonically increasing, which means that $p_x(x) \geq 0$.
2. $P(x) = \int_{-\infty}^{x} p_x(x') dx'$, which means that $\int_{-\infty}^{\infty} p_x(x) = 1$, where $P(-\infty) = 0$ and $P(\infty) = 1$.
3. $P(x_2) - P(x_1) = \int_{x_1}^{x_2} p_x(x) dx$, which means that $\text{Prob}(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} p_x(x) dx$. 

Figure 3.1: Measuring a random variable $N$ times, $N \to \infty$, to form an ensemble gives the probability density function.
3.2.2 Mean, Variance, Moments

The expectation or mean value of $x$ is defined as

$$E\{x\} = \int_{-\infty}^{\infty} x p_x(x) \, dx = m_1,$$  \hspace{1cm} (3.10)

whereas the parameter $\sigma^2$ is called the variance and is defined as

$$\sigma^2 = \int_{-\infty}^{\infty} (x - m_1)^2 p_x(x) \, dx,$$  \hspace{1cm} (3.11)

where $\sigma$ is called the standard deviation of $x$. In general, we can define an expectation value of a function of $x$ by the integral

$$\langle f(x) \rangle = E_x\{f(x)\} = \int_{-\infty}^{\infty} f(x) p_x(x) \, dx.$$  \hspace{1cm} (3.12)

For example, $f(x) = x$ gives the mean, $f(x) = (x - m_1)^2$ gives the variance, $f(x) = x^2$ gives the mean squared, $f(x) = x^n$ gives the $n$th moment $m_n$ of $x$, and $f(x) = (x - m_1)^n$ defines the $n$th central moment $\mu_n$ of $x$. In particular, the first three central moments have the properties that the zeroth central moment is given by

$$\mu_0 \equiv m_0 = 1,$$  \hspace{1cm} (3.13)

and the first central moment is defined by

$$\mu_1 = \int (x - m_1) p_x(x) \, dx,$$  \hspace{1cm} (3.14)

which can also be expressed as

$$\mu_1 = \int x p_x(x) \, dx - m_1 \int p_x(x) \, dx.$$  \hspace{1cm} (3.15)

But as the first integral is just $m_1$ and the second one is normalized to one, then $\mu_1$ is given by

$$\mu_1 = 0,$$  \hspace{1cm} (3.16)
Figure 3.3: A Gaussian (normal) probability density function, given by
\[ p_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x-\langle x \rangle)^2}{2\sigma^2} \right\}, \]
where \( \langle x \rangle \equiv m_1 \).

and the second central moment is defined by
\[ \mu_2 = \int (x - m_1)^2 p_x(x) \, dx, \tag{3.17} \]
which, as mentioned above, is often called \( \sigma^2 \). We can express \( \sigma^2 \) in terms of the moments \( m \) by expanding out the square to obtain
\[ \mu_2 = \int (x^2 - 2xm_1 + m_1^2) p_x(x) \, dx, \tag{3.18} \]
which can be simplified to
\[ \mu_2 = \int x^2 p_x(x) \, dx - 2m_1 \int xp_x(x) \, dx + m_1^2 \int p_x(x) \, dx. \tag{3.19} \]
The first integral is just \( m_2 \), the second integral is \( m_1 \), and the third is normalized to unity such that
\[ \mu_2 = m_2 - m_1^2. \tag{3.20} \]
If the function \( p_x(x) \) is considered a mass density along the \( x \)-axis, then \( m_1 \) would be its center of gravity, \( m_2 \) its moment of inertia, \( \sigma^2 \) its central moment of inertia, and \( \sigma \) the radius of gyration.

The constants \( m_1 \) and \( \sigma \) give only a limited description of \( p_x(x) \). Other moments give additional information; for example, if two densities have the same \( m_1 \) and \( \sigma \), their other moments can still be quite different. However, an important distribution, the Gaussian, is uniquely determined by moments through the second order, as we will see in the next paragraph. In all cases, if \( m_n \) is known for all \( n \), \( p_x(x) \) is determined uniquely. We will see how to explicitly construct \( p_x(x) \) from its moments after we have introduced the characteristic function.

The moments of a random variable are not arbitrary numbers—they have to satisfy some inequalities; for example,
\[ \sigma^2 = m_2 - m_1^2 \geq 0. \tag{3.21} \]
Similarly, since \( E \{(x^n - a)^2\} = m_{2n} - 2am_n + a^2 \) is non-negative for all values of \( a \), we have that, when \( a = m_n \),
\[ m_{2n} \geq m_n^2. \tag{3.22} \]
Let us now consider a zero-mean (i.e. $\langle x \rangle = 0$) Gaussian probability density function (see Figure 3.3 for a non-zero mean example which could be converted to zero mean by setting $m_1 = 0$),

$$p_x(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-x^2/2\sigma^2},$$

and look at its moments:

$$E\{x^n\} = \left\{\begin{array}{ll}
0, & n = 2k + 1 \\
(1)(3)(5) \cdots (n-1)\sigma^n, & n = 2k.
\end{array}\right.$$  \hspace{1cm} (3.24)

The odd moments are equal to zero because $p_x(x)$ is an even function, so that

$$p_x(x) = p_x(-x),$$

and therefore the odd moments can be written as

$$\int_{-\infty}^{\infty} x^{2n+1} p_x(x) \, dx = \int_{-\infty}^{0} |x|^{2n+1} p_x(x) \, dx + \int_{0}^{\infty} |x|^{2n+1} p_x(x) \, dx = 0.$$  \hspace{1cm} (3.26)

To derive the even moments of (3.24) above, we differentiate with respect to $\alpha$ $2k$ times in the equation

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} \, dx = \frac{\sqrt{\pi}}{\alpha},$$

which gives

$$\int_{-\infty}^{\infty} x^{2k} e^{-\alpha x^2} \, dx = \frac{(1)(3) \cdots (2k-1)}{2^k} \frac{\sqrt{\pi}}{\alpha^{2k+1}}.$$  \hspace{1cm} (3.28)

Then using that

$$\alpha = \frac{1}{2\sigma^2}$$

gives that

$$\int_{-\infty}^{\infty} x^{2k} p_x(x) \, dx = (1)(3) \cdots (2k-1)\sigma^{2k}.$$  \hspace{1cm} (3.30)

We see clearly in this form that the Gaussian is uniquely specified once its mean and variance are known, as all other moments can be expressed in terms of these. For this zero-mean Gaussian probability density function, we see that

$$m_2 = \sigma^2$$

$$m_4 = 3\sigma^4 = 3m_2^2,$$

which clearly satisfies the inequality of (3.22).

Other commonly encountered continuous probability density functions are shown in Figure 3.4. The Gaussian density is encountered in thermal noise phenomena, the uniform density describes white noise, the delta density describes a deterministic process with a fixed value, the log normal density describes atmospheric wave propagation, the Rayleigh density is encountered in black-body radiation, the Rician density describes a monochromatic source with additive Gaussian noise, the $t$ density describes a density that approaches Gaussian when its count parameter $M$ becomes large, and the gamma (beta) density describes...
the intensity statistics of speckles in an open (closed) speckle pattern. In optical communications, due to the quantum nature of light and the Gaussian nature of thermal noise, we will have to consider discrete probability distribution functions of photon count values as well as continuous densities of current levels.

The discrete probability distribution we will most often encounter is the Poisson distribution, and in fact all of the discrete count distribution functions we will use in optical communications will correspond in general to “conditioned” versions of the Poisson distribution. A random variable is Poisson distributed with parameter \( m \) if it takes values 0, 1, 2, ..., \( n \) with

\[
\text{Prob}(x = k) = e^{-m} \frac{m^k}{k!}, \quad k = 0, 1, \ldots
\]  

(3.32)

Therefore, \( x \) has density

\[
p_x(x) = e^{-m} \sum_{k=0}^{\infty} \frac{m^k}{k!} \delta(x - k).
\]  

(3.33)

Rather than considering \( x \) as a function of \( k \) and therefore considering a density \( p_x(x) \), we will simply use the notation \( p_k(k) \) for a count distribution and omit the delta function that shows up in (3.33). The distribution and density functions for a Poisson variable are shown in Figure 3.5. The mean as well as the variance of a Poisson distribution are equal to \( m \). We can see this if we first write

\[
\langle k \rangle = \sum k p_k(k),
\]  

(3.34)

where we will use the angular bracket notation for expectation values, and then expand out:

\[
\langle k \rangle = e^{-m} \sum \frac{km^k}{k!}.
\]  

(3.35)

Cancelling one of the \( k \)'s gives

\[
\langle k \rangle = e^{-m} \sum \frac{m^k}{(k-1)!},
\]  

(3.36)

which we can also write as

\[
\langle k \rangle = e^{-m} m \sum \frac{m^k}{k!}
\]  

(3.37)

by changing the summing index. The sum now is just the exponential, and therefore we can write

\[
\langle k \rangle = m.
\]  

(3.38)

A similar argument can be used to find

\[
\langle k^2 \rangle = m^2 + m
\]  

(3.39)

and thereby

\[
\langle k^2 \rangle - \langle k \rangle^2 = m.
\]  

(3.40)

The density that will take on primary importance in further considerations, as discussed above, is the conditional Poisson distribution, which can be expressed as

\[
p_k(k) = \int dm \text{pos}(k,m)p_m(m),
\]  

(3.41)

where

\[
\text{pos}(k,m) = e^{-m} \frac{m^k}{k!}
\]  

(3.42)
Figure 3.4: Sketches of (a) the uniform, (b) the delta, (c) the lognormal, (d) the Rayleigh, (e) the Rician, (f) the $t$, (g) the gamma, and (h) the beta distributions. The $I_m$ is the $m^{th}$-order imaginary Bessel function.
and where the $p_m(m)$ is a conditioning density. For a laser source biased well above threshold, one could say that the $m$ is well-defined and write

$$p_m(m) = \delta(m - \bar{m}_s)$$

(3.43)

to find that

$$p_k(k) = e^{-\bar{m}_s} \frac{\bar{m}_s^k}{k!},$$

(3.44)

which is again the Poisson distribution. Conversely, one could think of a thermal source. For a single temporal mode (to be defined soon during the discussion of the Karhunen-Loeve expansion) of a thermal source, the $p_m(m)$ is given by

$$p_m(m) = \frac{1}{\bar{m}_s} e^{-m/\bar{m}_s}.$$ 

(3.45)

Assuming that there are $N$ temporal modes in the conditioning period, one finds that the count statistics become

$$p_k(k) = \binom{N + k - 1}{k} \left( \frac{1}{1 + \bar{m}_s} \right)^N \left( \frac{\bar{m}_s}{1 + \bar{m}_s} \right)^k,$$

(3.46)

where the binomial coefficients are defined by

$$\binom{N}{k} = \frac{N!}{(N-k)!k!}.$$ 

(3.47)

This density is known as the negative binomial random variate, a generalization of the binomial distribution which is usually written as

$$p_k(k) = \binom{n}{k} p^k (1-p)^{n-k},$$

(3.48)

where the $k$ value varies from 0 to $n$. What we have here we note is of the form

$$p_k(k) = \binom{n + k - 1}{k} p^k (1-p)^n,$$

(3.49)
and therefore \( k \) can take on all values, not just values less than \( n \). A last interesting density is that generated from adding white Gaussian noise to a deterministic signal. If \( N \) temporal Gaussian modes are added to the laser level, then one obtains the \( N \)-times convolved Laguerre distribution

\[
p_k(k) = \frac{\bar{m}_i^k}{(1 + \bar{m}_i)^{k+N+1}} \exp \left\{ \frac{-\bar{m}_c}{\bar{m}_i(1 + \bar{m}_i)} \right\} L_k^N \left[ \frac{-\bar{m}_c}{\bar{m}_i(1 + \bar{m}_i)} \right], \tag{3.50}
\]

where \( \bar{m}_i \) is the incoherent conditioning parameter, \( \bar{m}_c \) is the coherent conditioning parameter, and \( L_k^N \) is the \( N \)-th-order Laguerre polynomial defined by

\[
L_k^N(x) = \sum_{\ell=0}^{k-N} (-1)^\ell \binom{k+N}{k-N} \frac{x^\ell}{\ell!}, \tag{3.51}
\]

which is the \((N+1)\)-times self-convolved Laguerre density. It should be noted at this point that the Poisson and the negative binomial random variate densities will be the primary ones used for optical count statistics which are received from a guided wave channel. The \((N+1)\)-times convolved Laguerre density is one that is often used for a free-space channel corrupted by background thermal radiation.

### 3.2.3 The Characteristic Function

The characteristic function of a random variable \( x \) with a probability density function \( p_x(x) \) is defined by the integral

\[
\psi_x(\omega) = \int_{-\infty}^{\infty} p_x(x) e^{j\omega x} \, dx = E\{e^{j\omega x}\}. \tag{3.52}
\]

Note that this is the Fourier transform of \( p_x(x) \). The characteristic function is also called the moment-generating function of \( x \), because if it is known, all the moments of \( x \) can be found. Note that one can express the expectation of \( x \) by

\[
E_x(x) = E\left\{ \frac{1}{j} \frac{\partial}{\partial \omega} \left[ e^{j\omega x} \right]_{\omega=0} \right\}, \tag{3.53}
\]

which can further be written as

\[
E_x(x) = \frac{1}{j} \frac{\partial}{\partial \omega} \left[ E\{e^{j\omega x}\} \right]_{\omega=0}, \tag{3.54}
\]

which can be expressed in terms of the characteristic function by

\[
E_x(x) = \frac{1}{j} \frac{\partial}{\partial \omega} \left[ \psi_x(\omega) \right]_{\omega=0}. \tag{3.55}
\]

After differentiating \( \psi_x(\omega) \) \( n \) times, we find

\[
\psi_x^{(n)}(\omega) = E\{x^n e^{j\omega x}\} = j^n \bar{m}_n
\]

\[
\psi_x^{(n)}(\omega = 0) = E\{x^n\} = j_n \bar{m}_n
\]

\[
\psi_x(\omega) = \psi_x(0) + \omega \frac{\partial}{\partial \omega} \psi_x(\omega = 0) + \cdots = \sum_n \frac{(j\omega)^n}{n!} \bar{m}_n. \tag{3.56}
\]

This means that, if the \( m_n \) are known, we know \( \psi_x(\omega) \) and vice versa.

In the important case of the Gaussian, we can write that

\[
\psi_x(\omega) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\bar{m}_1)^2}{2\sigma^2}} e^{j\omega x} \, dx. \tag{3.57}
\]
Note that this integral can be solved by a technique known as completing the squares; that is, we know that
\[
\int_{-\infty}^{\infty} e^{-a^2 x^2} \, dx = \frac{1}{a} \sqrt{\pi}.
\] (3.58)

In addition, one can write that
\[
j\omega x - \frac{x^2}{2\sigma^2} + \frac{x m_1}{\sigma^2} = -\frac{1}{2\sigma^2} \left[ x - (m_1 + j\omega \sigma^2) \right]^2 + \frac{1}{2\sigma^2} (m_1 + j\omega \sigma^2)^2.
\] (3.59)

So we have
\[
\psi_x(\omega) = \frac{1}{\sigma \sqrt{2\pi}} e^{-m_1^2 / 2\sigma^2} e^{(m_1 + j\omega \sigma^2)^2 / 2\sigma^2} \int e^{-x^2 / 2\sigma^2} \, dx
\] (3.60)
and therefore finally
\[
\psi_x(\omega) = e^{j\omega m_1} e^{-\omega^2 \sigma^2 / 2}.
\] (3.61)

From here we see that, when \(m_1 = 0\), \(\psi_x(\omega) = e^{-\omega^2 \sigma^2 / 2}\). The form of (3.61) gives graphic proof to the above-made assertion that a Gaussian probability is uniquely determined by its first two moments alone.

The most interesting densities that we will consider in this book, as previously mentioned, are the Poisson and the two densities that arise from conditional Poisson counting, the negative binomial random variate and \(N\) times convolved Laguerre. We’ll now proceed to discuss these one at a time.

The characteristic function for the Poisson is defined as
\[
\psi_k(\omega) = E[e^{j\omega k}],
\] (3.62)
which can be expanded out to yield
\[
\psi_k(\omega) = \sum_k e^{j\omega k} p_k(k).
\] (3.63)

Using the definition of \(p_k(k)\),
\[
\psi_k(\omega) = \sum_k e^{j\omega k} e^{-m} \frac{m^k}{k!}.
\] (3.64)

Rewriting \(\psi_k(\omega)\) in the form
\[
\psi_k(\omega) = e^{-m} \sum_k \frac{(me^{j\omega})^k}{k!}
\] (3.65)
and using an identity for the definition of the exponential gives us
\[
\psi_k(\omega) = \exp \left\{ m(e^{j\omega} - 1) \right\}.
\] (3.66)

For a conditional Poisson density, one can define a characteristic function by
\[
\psi_k(\omega) = E[e^{j\omega k}] = \sum_k p_k(k)e^{j\omega k}
\] (3.67)
and then use the definition of expectation to obtain
\[
\psi_k(\omega) = \sum_k \int dm \, \text{pos}(k, m)p_m(m)e^{j\omega k}
\] (3.68)
and then write out the $p_k(k)$ to find
\[ \psi_k(\omega) = \int dm \sum_k e^{-m} \frac{m^k}{k!} p_m(m) e^{j\omega k}. \] (3.69)
The $p_m(m)$ is independent of $k$, so one can write
\[ \psi_k(\omega) = \int dm e^{-m} p_m(m) \sum_k \frac{(me^{j\omega})^k}{k!}. \] (3.70)

The summation is just an exponential, so rearranging terms leads to
\[ \psi_k(\omega) = \int dm p_m(m) \exp \{m(e^{j\omega} - 1)\}, \] (3.71)
which by definition is
\[ \psi_k(\omega) = \psi_m[-j(e^{j\omega} - 1)], \] (3.72)
which allows us to find the characteristic function for the count statistic directly from the characteristic function for the condition number $m$.

Let’s say that we have a random variable $k$ that is a sum over a number of other random variables $k_i$ such that
\[ k = \sum_i k_i. \] (3.73)
If the $k_i$ are independent, then we could write that
\[ \psi_k(\omega) = E[e^{j\omega k}]. \] (3.74)
Writing the sum for $k$ now explicitly, we have
\[ \psi_k(\omega) = E[e^{j\omega \sum k_i}]. \] (3.75)

We can write the exponentials as products to get
\[ \psi_k(\omega) = E_{k_i}[e^{j\omega k_i}]E_{k_2}[e^{j\omega k_2}] \cdots E_{k_n}[e^{j\omega k_n}], \] (3.76)
which can be written in more compact notation as
\[ \psi_k(\omega) = \prod_{i=1}^n \psi_{k_i}(\omega), \] (3.77)
or the characteristic function of the summed variables is the product of the characteristic function of the individual “modes” of the process. Inverse transforming the above result gives that
\[ p_k(k) = \int d\omega e^{-j\omega k} \psi_k(\omega). \] (3.78)
This is to say that, if the variables are independent and we know the individual characteristic functions, we can always at least find an expression for the density. Using the convolution theorem for the Fourier transform yields
\[ p_k(k) = p_{k_1}(k_1) * p_{k_2}(k_2) * \cdots * p_{k_n}(k_n), \] (3.79)
where the asterisk denotes convolution, defined by
\[ f(t) * g(t) = \int f(t-t')g(t')dt' = \int f(t')g(t-t') \, dt', \] (3.80)
which in our case, however, is the discrete version of the convolution, defined by

\[
p_{k_1}(k_1) \ast p_{k_2}(k_2) = \sum_{k'} p_{k_1}(k_1 - k')p_{k_2}(k') = \sum_{k'} p_{k_1}(k')p_{k_2}(k_2 - k').
\] (3.81)

The form of (3.79), though, is not nearly so nice as the form of (3.27), which is a prime reason for using characteristic functions in multivariate problems.

For a single temporal mode of a thermal light source, it was earlier noted that

\[
p_m(m) = 1\bar{m} e^{-m/\bar{m}}.
\] (3.82)

The characteristic function for this process can then be found to be

\[
\psi_m(\omega) = E[e^{j\omega m}] = \frac{1}{\bar{m}} \int e^{-m/\bar{m}} e^{j\omega m} dm = \frac{1}{1 - j\omega \bar{m}}.
\] (3.83)

If there were \(N\) modes radiated by the source, then one could write that

\[
\psi_m(\omega) = \left(\frac{1}{1 - j\omega \bar{m}}\right)^N.
\] (3.84)

The count distribution due to this source could then be expressed as

\[
p_k(k) = \int d\omega e^{-j\omega k} \psi_k(\omega).
\] (3.85)

Plugging in for the \(\psi_0(\omega)\), we find

\[
p_k(k) = \int d\omega e^{-j\omega k} \left[ -j(e^{j\omega} - 1) \right],
\] (3.86)

and using (3.84) above gives that

\[
p_k(k) = \int d\omega e^{-j\omega k} \left[ \frac{1}{1 - (e^{j\omega} - 1)\bar{m}} \right]^N.
\] (3.87)

The integral is not hard to perform if one notes that the denominator has a simple pole at

\[
\omega = j \ln \left(\frac{m}{m + 1}\right).
\] (3.88)

Application of Cauchy’s residue theorem for an \(n\)th-order pole then allows one to find the negative binomial random variate density of equation (3.46).

In the case where thermal noise is added to a fixed-value coherent field, one can show that the level density and its characteristic function for a single temporal mode are given by

\[
p_m(m) = \frac{1}{\bar{m}^1} \exp \left\{ -\frac{m + \bar{m}}{\bar{m}} \right\} I_0 \left[ \frac{m\bar{m}}{\bar{m}_1} \right]
\]

\[
\psi_m(\omega) = \frac{\exp \left[ \frac{m\bar{m}_1}{m_1 - j\omega} \right]}{1 - \bar{m}_1 j\omega},
\] (3.89)
where $\bar{m}_i$ is the magnitude of the thermal noise field, $\bar{m}_c$ is the coherent field, and $I_0(x)$ is the imaginary Bessel function of argument $x$ defined by

$$I_0(x) = \sum_{j=0}^\infty \frac{(x/2)^{2j}}{(j!)^2}. \quad (3.90)$$

In this case, the characteristic function, to the best of my knowledge, is not obtainable in closed form, and different techniques than those used in the preceding examples must be used. Despite the fact that this is a subsection in characteristic functions, we will proceed to answer this issue. The count statistics for this single-mode case can be found from

$$p_k(k) = \int dm \text{pos}(k,m)p_m(m), \quad (3.91)$$

where the integral can be performed to give

$$p_k(k) = \text{Lag}[k,\bar{m}_c,\bar{m}_i,0], \quad (3.92)$$

where the Laguerre counting probability function is defined by

$$\text{Lag}[k, a, b, c] = \left[\frac{b^k}{1+b^{k+c+1}}\right] \exp\left\{-a\frac{1+b}{1+b}\right\} L_k^c \left[\frac{-a}{b(1+b)}\right]. \quad (3.93)$$

For $N$ modes, one must convolve this density with itself $N$ times to obtain

$$p_k(k) = \frac{\bar{m}_i}{(1+\bar{m}_i)^{k+N+1}} \exp\left\{-\frac{\bar{m}_c}{1+\bar{m}_i}\right\} L_k^N \left[\frac{-\bar{m}_c}{\bar{m}_i(1+\bar{m}_i)}\right]. \quad (3.95)$$

### 3.2.4 Transformation of Variables

What if we know $p_x(x)$ and we wish to find $p_y(y)$, where $y = f(x)$ is some function of $x$? The probability density function of $y$ is given by

$$p_y(y) = \left. \frac{p_x(x)}{\left| \frac{\partial f}{\partial x} \right|} \right|_{x=f^{-1}(y)}. \quad (3.96)$$

Let’s say for example that we have a density $p_n(n)$ and we want to transform to a density $p_m(m)$ where $m$ is defined by

$$m = n \tau_d. \quad (3.97)$$

We then could write

$$p_m(m) = \left. \frac{p_n(n)}{\partial(n\tau_d)/\partial n} \right|_{n=m/\tau_d}, \quad (3.98)$$

which can be evaluated to give

$$p_m(m) = \frac{1}{\tau_d} p_{m/\tau_d}(m/\tau_d). \quad (3.99)$$

For example, say that

$$p_n(n) = \frac{1}{\bar{n}} e^{-n/\bar{n}}. \quad (3.100)$$

Then

$$p_m(m) = \frac{1}{\bar{m}} e^{-m/\bar{m}}. \quad (3.101)$$
3.3 Joint Random Variables

3.3.1 Definition

As an example of joint random variables, consider a laser. In a laser the phase and amplitude of emitted light are joint random variables. In a gas laser, however, they follow different statistics than in a semiconductor laser. In gas lasers, the phase and amplitude are uncorrelated, since all the molecules are independent sources. In a semiconductor laser, the two quantities are correlated due to band filling. That is, changes in amplitude change the inversion level, which in turn changes the transition frequency, which causes phase transients. Therefore we can consider the phase and amplitude of the semiconductor laser to be joint random variables. Certainly there are many other examples, but lasers are important to us. In general, the joint distribution
\[ P_{xy}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{xy}(x, y) \, dx \, dy \]
gives the probability of the event for which \( x \) and \( y \) fall into intervals \( I_x \) and \( I_y \) in the \( xy \) plane:

- \( \text{Prob}[x \in I_x, y \in I_y] = \int_{I_x} \int_{I_y} p_{xy}(x, y) \, dx \, dy \)
- \( \text{Prob}[x_1 < x < x_2, y_1 < y \leq y_2] = P(x_2, y_2) - P(x_1, y_1) \)
- \( \text{Prob}[x < x_1, y_1 < y < y_2] = P(x, y_2) - P(x, y_1) \)
- \( \text{Prob}[x_1 < x \leq x_2, y_1 < y \leq y_2] = P(x_2, y_2) - P(x_1, y_2) - P(x_2, y_1) + P(x_1, y_1) \). (3.103)

In the case of several random variables that describe an event, the statistics of each are called marginal. So, \( p_x(x) \) is the marginal density of \( x \), and \( p_y(y) \) is the marginal density of \( y \). The marginal statistics of \( x \) and \( y \) can be expressed in terms of their joint function \( p_{xy}(x, y) \) as follows:

- \( p_x(x) = \int_{-\infty}^{\infty} p_{xy}(x, y) \, dy \)
- \( p_y(y) = \int_{-\infty}^{\infty} p_{xy}(x, y) \, dx \). (3.104)
If the random variables \( x \) and \( y \) are independent, then
\[ p_{xy}(x, y) = p_x(x)p_y(y). \] (3.105)

If \( x + y = z \), we have
\[ p_z(z) = \int_{-\infty}^{\infty} p_{xy}(z-y, y) dy = \int_{-\infty}^{\infty} p_x(z-y)p_y(y) dy = p_x*p_y. \] (3.106)

The above integral is again the convolution of the functions \( p_x(x) \) and \( p_y(y) \), which we will often denote by the asterisk as we did before in equation (3.80). So, if two random variables are independent, then the density of their sum equals the convolution of their densities. For example, suppose that resistors \( R_1 \) and \( R_2 \) are two independent random variables uniform between 900 and 1000 \( \Omega \). Then, if they are connected in series, the density of the resulting resistor \( R = R_1 + R_2 \) is a triangle between 1800 and 2000 \( \Omega \).

### 3.3.2 Joint Moments

Given two random variables \( x \) and \( y \) and a function \( f(x, y) \), we can form the random variable \( z = f(x, y) \). The expected value of \( z \) is given by
\[ E\{z\} = E\{f(x, y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)p_{xy}(x, y) dx dy. \] (3.107)

If the random variables \( x \) and \( y \) are discrete and the values \( x_i \) and \( y_k \) have a probability \( p_{ik} \), then
\[ E\{f(x, y)\} = \sum_{i,k} f(x_i, y_k)p_{ik}. \] (3.108)

Linearity follows from the above:
\[ E\left\{\sum_{k=1}^{n} a_k f_k(x, y)\right\} = \sum_{k=1}^{n} a_k E\{f_k(x, y)\}. \] (3.109)

So, \( E\{x + y\} = E\{x\} + E\{y\} \), while it is not always true that \( E\{xy\} \neq E\{x\}E\{y\} \).

The covariance \( C_{xy} \) of two random variables \( x \) and \( y \) is by definition
\[ C_{xy} = E\{(x - m_x)(y - m_y)\}, \] (3.110)
where \( E\{x\} = m_x \) and \( E\{y\} = m_y \). Expanding the previous equation gives
\[ C_{xy} = E\{xy + m_xm_y - ym_x - xm_y\} = E\{xy\} - E\{x\}E\{y\}. \] (3.111)

The correlation coefficient \( R_{xy} \) of the random variables \( x \) and \( y \) is by definition
\[ R_{xy} = \frac{C_{xy}}{\sigma_x\sigma_y} \] (3.112)

Also, \(|R| \leq 1 \) and \(|C_y| \leq \sigma_x\sigma_y \). This can be seen from
\[ E\{[a(x - m_x) + (y - m_y)]^2\} = a^2\sigma_x^2 + 2aC_{xy} + \sigma_y^2. \] (3.113)

This is a positive quadratic, so its discriminant is negative:
\[ C_{xy}^2 - \sigma_x^2\sigma_y^2 \leq 0. \] (3.114)
Two random variables are uncorrelated if $C_{xy} = 0$ and $R_{xy} = 0$. Also, in this case $E\{xy\} = E\{x\}E\{y\}$. Two random variables are orthogonal if

$$E\{xy\} = 0.$$  \hspace{1cm} (3.115)

For example, if $x$ and $y$ are uncorrelated, then $(x - m_{x1})$ and $(y - m_{y1})$ are orthogonal. If, further, $m_{x1} = m_{y1} = 0$, then $x$ and $y$ are orthogonal.

If two random variables are independent,

$$p_{xy}(x, y) = p_x(x)p_y(y),$$  \hspace{1cm} (3.116)

then they are also uncorrelated as

$$E\{xy\} = \int\int xyp_x(x)p_y(y) \, dx \, dy = E\{x\}E\{y\}.$$  \hspace{1cm} (3.117)

The mean of the product $x^p y^q$ is by definition the joint moment of $x$ and $y$ of order $p + q = n$:

$$M_{10} = m_{x1} \hspace{1cm} M_{01} = m_{y1} \hspace{1cm} \text{first-order moments}$$

$$M_{20} = E\{x^2\} \hspace{1cm} M_{11} = E\{x\}E\{y\} \hspace{1cm} M_{02} = E\{y^2\} \hspace{1cm} \text{second-order moments}.$$  \hspace{1cm} (3.119)

### 3.3.3 Joint Characteristic Function

The joint characteristic function is the integral

$$\Phi(\omega_x, \omega_y) = \int\int p_{xy}(x, y)e^{j(\omega_x x + \omega_y y)} \, dx \, dy.$$  \hspace{1cm} (3.120)

We can see that

$$\Phi_{xy}(\omega_x, \omega_y) = E\{e^{j(\omega_x x + \omega_y y)}\}.$$  \hspace{1cm} (3.121)

If $z = ax + by$, then

$$\Phi_z(\omega) = E\{e^{j(ax+by)\omega}\} = \Phi_{xy}(a\omega, b\omega) = \Phi_{xy}(\omega_x, \omega_y).$$  \hspace{1cm} (3.122)

### 3.4 Sequences of Random Variables and Stochastic Processes

Let us consider a random process $x$ that is also a function of the time $t$, such as is depicted in Figure 3.7. Indeed, Figure 3.7 represents the situation which we actually have, that in which we receive a signal as a function of time, and that signal has both a deterministic information portion as well as a noise portion. The noise portion of the received signal contains a wealth of information about the world in general but has nothing specific to add to what the sender is trying to tell us and in fact tends to mask what the sender is trying to tell us. (The situation is reversed in Chapter 14 on sensing.) The question is how to handle such a signal in terms of the formulation we have been following for the last eighty or so equations.

Clearly, we can think of a stochastic function of time (stochastic process) as the limit of a discrete set of events as is depicted in Figure 3.7. In this sense, we can think of the density $p_{v(t)}(v(t))$ of this process as a limit:

$$p_{v(t)}(v(t)) = \lim_{\Delta t \to 0} p_v(V),$$  \hspace{1cm} (3.123)
where

\[
V = \begin{bmatrix}
V_1 \\
V_2 \\
\vdots \\
V_N
\end{bmatrix} = \begin{bmatrix}
v(t_1) \\
v(t_2) \\
\vdots \\
v(t_N)
\end{bmatrix}
\] (3.124)

and

\[t_{j+1} - t_j = \Delta t.\] (3.125)

One can consider the \(p_v(V)\) in this case as a joint density of the \(v(t_i) = V_i\) with the corresponding definition that

\[E\{f(V)\} = \int p_v(v) dV_1 dV_2 \cdots dV_n.\] (3.126)

When we talked about joint densities, an important concept was that of independence. Were the process \(v(t)\) sufficiently random and the \(\Delta t\) sufficiently long, the \(V_i\) might become independent, which would mean that the joint density function \(p_v(v)\) could be written in the form

\[p_v(v) = p_{v_1}(V_1)p_{v_2}(V_2) \cdots p_{v_n}(V_n).\] (3.127)

Equation (3.91) means that, were we to take expectations of products of \(V_i\) such as the product of all the \(V_i\)'s, we could express the result as

\[E[V_1 V_2 \cdots V_n] = E[V_1]E[V_2] \cdots E[V_n].\] (3.128)

As there was a product of \(n\) \(V_i\)'s in the argument of the expectation, we would say that the expectation was an \(n\)-th-order statistic. In particular, we often define the second-order, centered expectation to be the correlation function \(R(V_i, V_j)\):

\[R_{V_i V_j}(V_i, V_j) = E[(V_i - m(V_i))(V_j - m(V_j))].\] (3.129)

The matrix elements

\[M_{ij} = R_{V_i V_j}(V_i, V_j),\] (3.130)

taken collectively, are often called the elements of the covariance matrix. Clearly, for independent \(V_i\), one can write that

\[R_{V_i V_j}(V_i, V_j) = E[(V_i - m(V_i))(V_j - m(V_j))] = E[(V_j - m(V_i))^2] \delta_{ij},\] (3.131)
where $\delta_{ij}$ is the Kronecker delta function.

In general, the vector of $V_i$’s may have correlations of all orders. Even in the case where the values are so-called uncorrelated as in (3.131), there may be higher-order correlations. There is an important class of joint random variables for which second-order statistics are sufficient. Consider the joint density

$$p_V(v) = \frac{1}{\sqrt{2\pi}^{N/2} \sqrt{\det M}} e^{-\frac{1}{2} v^T M^{-1} v},$$

(3.132)

where $M$ is the covariance matrix of (3.130) and $v$ is defined by

$$v = \begin{bmatrix} V_1 - E(V_1) \\ \vdots \\ V_N - E(V_N) \end{bmatrix}.$$  

(3.133)

This is the joint Gaussian density. It is quite easy to show that when these variables become uncorrelated they also become independent. That this Gaussian case is so important is due to a couple of facts. One is the law of large numbers. This law tells us that, if we have a process which is due to a multitude of other processes such that

$$T = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i,$$

(3.134)

then the main peak of the density function of $T$ will approach a Gaussian in the limit of large $N$ independently of the density functions of the $x_i$. It is reasonably easy to prove this assertion, and we will use an approach borrowed from Garrod (1995). Consider $N$ independent variables $x_1, x_2, \ldots, x_N$, each with a distribution $p_n(x_n)$. Without any real loss of generality, we will assume that all of these (assumed real) variables have zero mean. Let’s then define the quantity $T$ by

$$T = \lambda \sum_{n=0}^{N} x_n,$$

(3.135)

where we will eventually let $\lambda$ be the inverse square root of $N$. We can now define the characteristic function of the $T$ by

$$\psi_T(\omega) = \prod_{n=1}^{N} \psi_n(\lambda \omega),$$

(3.136)

where we have used an earlier argument in this chapter to allow us to express the characteristic function of a sum of variables as a product of the characteristic functions of their individual distributions. A fundamental theorem in calculus states that

$$|e^{i\theta} - 1 - i\theta + \frac{\theta^2}{2}| \equiv R(\theta) < c|\theta|^3$$

(3.137)

for some value of $c$. If we then write the definition of the $n^{th}$ characteristic function,

$$\varphi_n(\lambda \omega) = \int e^{i\lambda \omega x_n} p_n(x_n) \, dx_n,$$

(3.138)

we can use the above theorem to approximate

$$\psi_n(\lambda \omega) \approx \int \left(1 + i\lambda \omega x_n - \frac{\lambda^2 \omega^2 x_n^2}{2}\right) p_n(x_n) \, dx_n,$$

(3.139)
but we immediately see that
\[ \int p_n(x_n) \, dx_n = 1 \]
\[ \int x_n p_n(x_n) \, dx_n = 0 \]
\[ \int x_n^2 p_n(x_n) \, dx_n = (\Delta x_n)^2, \] (3.140)
and we can therefore write that
\[ \psi_n(\lambda \omega) \approx 1 - \frac{\omega^2 \lambda^2}{2} (\Delta x_n)^2. \] (3.141)
We can then write that
\[ \ln \psi_T(\omega) = \ln \prod_{i=1}^{N} \psi_n(\lambda \omega), \] (3.142)
which can then be re-expressed in the form
\[ \ln \psi_T(\omega) = \sum_{i=1}^{N} \ln \left[ 1 - \frac{\omega^2 \lambda^2}{2N} (\Delta x_n)^2 \right]. \] (3.143)
We can now take
\[ \lambda = \frac{1}{\sqrt{N}} \] (3.144)
to rewrite the above as
\[ \ln \psi_T(\omega) = \sum_{i=1}^{N} \ln \left[ 1 - \frac{\omega^2}{2N} (\Delta x_n)^2 \right]. \] (3.145)
For a large enough \( N \), one would expect the second term in the brackets to be small. Using
\[ \ln(1 + \epsilon) \approx \epsilon \] (3.146)
for small enough \( \epsilon \), we can then write that
\[ \ln \psi_T(\omega) = -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\omega^2}{2} (\Delta x_n)^2. \] (3.147)
Then, assuming that \( 1/N \) times the sum exists, we write
\[ \frac{1}{N} \frac{\omega^2}{2} \sum_{i=1}^{N} (\Delta x_n)^2 = \frac{\omega^2 a^2}{2}, \] (3.148)
allowing us to write that
\[ \ln \psi_T(\omega) = -\frac{1}{2} \omega^2 a^2 \] (3.149)
or that
\[ \psi_T(\omega) = e^{-\frac{\omega^2 a^2}{2}}. \] (3.150)
The distribution of $T$ will then be given by

$$p_T(T) = \frac{1}{2\pi} \int e^{-\frac{\omega^2}{2a^2}} e^{-\omega T} d\omega.$$  

(3.151)

Noting that

$$a^2 \left[ \omega^2 + 2i \frac{\omega T}{a^2} \right] = a^2 \left( \omega + \frac{\omega T}{a^2} \right)^2 + \frac{T^2}{2a^2},$$  

(3.152)

we note that

$$p_T(T) = \frac{e^{-T^2/2a^2}}{2\pi} \int e^{-\frac{\omega^2}{2a^2}} \left( \omega + \frac{\omega T}{a^2} \right) d\omega,$$  

(3.153)

but the Gaussian integral is just given by

$$\int e^{-\frac{\omega^2}{2a^2}} \left( \omega + \frac{\omega T}{a^2} \right) d\omega = \frac{\sqrt{2\pi}}{a},$$  

(3.154)

and therefore we see that

$$p_T(T) = \frac{e^{-T^2/2a^2}}{\sqrt{2\pi}a},$$  

(3.155)

which is indeed a Gaussian. However, the terms that were ignored will cause the wings of the $T$ distribution to exponentially diverge from the Gaussian form.

It should be noted here that a corollary to the law of large numbers is the central limit theorem. (See, for example, Feller 1968, Chapter 10.) In its most general form, the theorem states that, if one samples a normal process $x$ (Gaussianly distributed random variable of mean $\mu$ and variance $\sigma^2$) $N$ times, the mean value of the $N$th sample, defined as

$$\bar{x} = \sum \frac{x_i}{N},$$  

(3.156)

where $x_i$ are the samples and $\bar{x}$ the mean, will differ from the actual mean (asymptotically with increasing $N$) by no more than the standard deviation $\sigma$ divided by the square root of $N$, or

$$\lim_{N \to \infty} (\bar{x} - \mu)^2 \approx \frac{\sigma^2}{N}.$$  

(3.157)

The central limit theorem is a central item in the theory of sampling. If one can look at a process more than once, one can lower the error probability (that is to say, the standard deviation from the mean) by the inverse square root of the number of samples. We will not prove this theorem, but we will find reason to invoke it in Chapter 12, section 2, where we discuss Monte-Carlo sampling, as well as throughout Chapter 14, where we discuss optical sensing.

This central limit theorem states essentially that a complicated enough process will always be a Gaussian in its main peak close enough to the mean value of the distribution (out to two or three $\sigma$'s). More particularly in our case, though, the Gaussian distribution is important because of the physics it represents. Thermal processes are due to a multitude of subprocesses and are therefore Gaussian in nature. Thermal light emission, thermal carrier excitation, and thermal circuit noise are all Gaussian processes. In fact, as this book progresses, one will note that, to deal with phenomena in optical communications, one needs to deal essentially only with Poisson and Gaussian processes and with the processes which can be formed from the interaction variables distributed according to the statistics of these two types of processes. In fact, the Poisson distribution even becomes Gaussian in the mean for large enough average received signals, albeit for a very different reason than that we can use a Gaussian to represent thermal processes, but we will see much more on this in Chapter 7. To carry out a demonstration of the Gaussian limit of the Poisson distribution, we need only recall the Stirling approximation that

$$\ln k! = \frac{1}{2} \ln 2\pi + \left( k + \frac{1}{2} \right) \ln k - k + \frac{1}{12}\frac{1}{k},$$  

(3.158)
which for large enough \( k \) reduces to
\[
\ln k! = k \ln k - k. \tag{3.159}
\]

The Poisson distribution is given by
\[
p_k(k) = \frac{m^k}{k!} e^{-m}, \tag{3.160}
\]
and therefore we can write that
\[
\ln p_k(k) = k \ln m - m - k \ln k + k, \tag{3.161}
\]
which we can rewrite as
\[
\ln p_k(k) = -k \ln \frac{k}{m} + k - m. \tag{3.162}
\]

Writing that
\[
\frac{k}{m} = 1 + \left( \frac{k - m}{m} \right), \tag{3.163}
\]
we can rewrite the logarithm as
\[
\ln p_k(k) = -m \left( 1 + \frac{k - m}{m} \right) \ln \left( 1 + \frac{k - m}{m} \right) + (k - m). \tag{3.164}
\]

If we now assume that
\[
\left| \frac{k - m}{m} \right|^3 << \left( \frac{k - m}{m} \right)^2, \tag{3.165}
\]
which is equivalent to
\[
\left| \frac{k - m}{m} \right| << 1, \tag{3.166}
\]
we can expand the right-hand side logarithm using the expansion
\[
\ln(1 + x) = x - \frac{x^2}{2} + \cdots \tag{3.167}
\]
to obtain
\[
\ln p_k(k) = -m \left( 1 + \frac{k - m}{m} \right) \left( \frac{k - m}{m} \right) \left( 1 - \frac{k - m}{2m} \right) + (k - m). \tag{3.168}
\]

Again expanding through second order, we see
\[
\ln p_k(k) = -\frac{(k - m)^2}{2m}, \tag{3.169}
\]
which gives the desired result:
\[
p_k(k) = Ne^{-(k-m)^2/2m}, \tag{3.170}
\]
where \( N \) is the normalization we have approximated away but for which we can solve using
\[
\int dk p_k(k) = 1, \tag{3.171}
\]
where we have now assumed \( k \) and \( m \) so large that \( k \) is effectively continuous, giving
\[
p_k(k) = \frac{1}{\sqrt{2\pi m}} e^{-\frac{4k^2}{m^2}},
\]
which is a Gaussian with both mean and variance given by the count variable \( m \). Of course, though, the actual Poisson is asymmetric about its mean, and the Gaussian is symmetric about its mean, so clearly the left-out terms are important far enough from the mean that the values of the distribution become small relative to the peak value. This is essentially always the case in optical communications, where we are always trying to minimize the amount of power per information period. But now enough on Gaussians and back to random processes!

In the limit of \( \Delta t \to 0 \), equation (3.126) takes on the rather formidable form
\[
E_{v(t)}(v(t)) = \int_{\text{all functional realizations}} dv(t)p_{v(t)}(v(t))f(v(t)),
\]
where the integral in (3.173) is Feynman’s integral over all paths (Feynman and Hibbs, 1965) and is in fact the form originally studied by Norbert Wiener in the late 1930s and one that is presently used extensively in statistical mechanics, quantum field theory, and condensed matter physics. Despite being well defined, the techniques necessary to evaluate (3.173) as an integral over all paths would take us too far afield. (Interestingly enough, the integral can be evaluated analytically for certain multidimensional Gaussian distributions.) Further, a primary reason that this type of integral works so well for equilibrium statistical mechanics lies in the equilibrium. A process in equilibrium can be considered to be essentially an ergodic process. Our optical communication signals, however, will need to carry information. An information-bearing signal cannot be considered to have stationary increments in terms of the detector period, but this means that we can only really define up through second order statistics, as will be discussed in more depth in section 11.1 of Chapter 11.

What we did in the vector case above was to consider correlations of various orders. For example, to find the expectation value \( E(v(t)) \), one need only define a density function \( p_v(v) \) by
\[
E[v(t)] = \int v p_v(v) dv,
\]
where now \( v \) is a simple number evaluated at a time \( t \), not a function of the time \( t \). To really define the process \( v(t) \) will require statistics of all orders such that one can find such things as
\[
E[v(t_1)v(t_2)\cdots v(t_n)] = \int dv_1 \cdots dv_n p_{v_1 \cdots v_n}(v_1 \cdots v_n)v_1 \cdots v_n,
\]
which looks a lot like some of the vector cases above but has a different meaning. The \( t_i \) here are not fixed but can vary. In the vector case, we could argue about independence of \( v_i \). If our process \( v(t) \) is a continuous one, we can always pick a \( t_1 - t_2 \) small enough that \( v(t_1) \) cannot be independent of \( v(t_2) \). Oftentimes one takes the correlation function to be a delta function in order to simplify mathematical calculations. Clearly, a continuous function (as observable things are) has to have enough memory to remain a continuous function, meaning that \( R_{v_1 v_2}(t_1, t_2) \) can never really become a delta function. A delta function representation for the correlation can only be an idealization given some set of time constants. The density functions in this stochastic process case then will not become simple in their higher orders. However, we will see that we can do a lot with only second-order statistics and in fact have to when discussing signals that have at best stationary increments.

Let us now consider in some detail the correlation function
\[
R_{v_1 v_2}(t_1, t_2) = E[v(t_1)v(t_2)],
\]
where the \( t_1 \) and \( t_2 \) notation is a bit of a simplification from the notation of (3.131) above. It should be noted that, if we are considering periods of time in which our signal is in a single state, then the correlation
function will only depend on the difference of \( t_1 \) and \( t_2 \), not on their absolute values. For binary signals, this would mean times which did not include a switching. For analog signals, this would mean times smaller than the inverse of the highest frequency modulated onto the carrier. More will be said on this in section 11.1. Clearly, the maximum value of the correlation function will be obtained when \( t_1 = t_2 \). The value will fall off with increasing \(|t_1 - t_2|\). When the function reaches some small value at a value of \( t \approx \tau_{coh} = |t_1 - t_2| \), we call that value of \(|t_1 - t_2|\) the coherence time. This is the time at which the process has lost its memory of where it was before.

An important property of the correlation function is embodied in the Wiener-Khintchine theorem. As stated above, the statistics of a process can remain quite constant for some period (for example, an information period). During this period, the correlation function is not a function of the absolute time but only the difference. This is the limit in which we will generally work. In this limit, we can write

\[
R(\tau) = E[v(t + \tau)v(t)].
\]  

(3.177)

There is an important theorem which tells us that, in this limit, one can find the spectral density of the process \( v(t) \) from the relation

\[
S_v(\omega) = \int e^{-i\omega\tau} R(\tau) d\tau = F_{\omega\tau}[R(\tau)],
\]

(3.178)

where the last equality sign notes that the mathematical operation is that of a Fourier transform and where the \( \omega\tau \) subscript on \( F \) means that the transform is from \( \tau \)-space to \( \omega \)-space. (In most of what follows, we will simply write \( F \) without the subscript.) The relation of (3.178) is generally referred to as the Wiener-Khintchine theorem. We can also write the inverse relation of (3.178) to obtain

\[
R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_v(\omega)e^{i\omega\tau} d\omega.
\]

(3.179)

Although the above definition of coherence time is quite loose, one can rigorously define temporal modes whose shapes give us the coherence packets using a technique developed independently by Karhunen (1947) and Loeve (1955). One can define a set of eigenvalues \( \gamma_j \) and eigenvectors \( \varphi_j(t) \) by the equation

\[
\int_0^{\tau_d} R(t_1, t_2) \varphi_j(t_2) dt_2 = \gamma_j \varphi_j(t_1),
\]

(3.180)

where \( \tau_d \) is something like what will become our detector time. As the \( \varphi_j \) form a complete set on the time interval \((0, \tau_d)\), one can in particular write that

\[
v(t) = \sum_j a_j \varphi_j(t),
\]

(3.181)

where we will show that the energy residing in the \( j \)th temporal mode is \( \gamma_j \). The beauty of this Karhunen-Loeve expansion is, then, that these \( a_j \) are uncorrelated random variables. This is actually quite easy to show. To derive the Karhunen-Loeve equations, we use equation (3.181) together with the independence condition

\[
\langle a_i a_j^* \rangle = \gamma_j \delta_{ij}
\]

(3.182)

to obtain the equation

\[
\langle v(t_1)v^*(t_2) \rangle = \sum_{ij} \langle a_i a_j^* \rangle \varphi_i(t_1)\varphi_j^*(t_2),
\]

(3.183)

which we can rewrite using (3.176) (using complex \( v(t) \)) and (3.182) as

\[
R(t_1, t_2) = \sum \gamma_j \varphi_i(t_1)\varphi_j^*(t_2).
\]

(3.184)
If we multiply (3.184) by $\varphi_k(t_2)$ and integrate using the orthogonality relation that

$$\int \varphi_k(t_2)\varphi_j^*(t_2) dt_2 = \delta_{jk},$$

we find (3.180). To relate the eigenvalues to the energy absorbed during a detector time $\tau_d$, we multiply both sides of (3.180) by $\varphi_j(t_1)$ and integrate $dt_1$ to obtain

$$\gamma_j = \int_0^{\tau_d} R(t_1, t_2)\varphi_j(t_1)\varphi_j(t_2) dt_1 dt_2. \tag{3.186}$$

Recalling that we can expand any function $f(t)$ in a complete orthonormal set $\varphi_j(t)$, we note that

$$f(t) = \sum_j \int f(t')\varphi_j(t') dt'\varphi_j(t) \tag{3.187}$$

and, rewriting as

$$f(t) = \int f(t') \sum_j \varphi_j(t')\varphi_j(t) dt', \tag{3.188}$$

which we note is the same form as

$$f(t) = \int f(t')\delta(t - t') dt, \tag{3.189}$$

we see that the delta function is expandable as

$$\delta(t - t') = \sum_j \varphi_j(t')\varphi_j(t). \tag{3.190}$$

If we then sum the equation for $\gamma_j$ over $j$, we obtain

$$\sum_j \gamma_j = \int_0^{\tau_d} R(t_1, t_2) \sum_j \varphi_j(t_1)\varphi_j(t_2) dt_1 dt_2 \tag{3.191}$$

to obtain that

$$\sum_j \gamma_j = \int_0^{\tau_d} R(t_1, t_2) dt'. \tag{3.192}$$

The correlation function takes on its maximum value of $P(t)$ with equal arguments—that is,

$$R(t, t) = P(t), \tag{3.193}$$

giving that

$$\sum_j \gamma_j = \int_0^{\tau_d} P(t') dt' \tag{3.194}$$
or, if the power were constant in a detector period,

$$\sum_j \gamma_j = P\tau_d. \tag{3.195}$$
With constant or not, the two above results indicate that the sum of eigenvalues is identical to the energy absorbed by the detector within a detector period $\tau_d$. This result also indicates to us that we cannot have an infinite number of energy-carrying modes but that instead only a finite number of modes have nonzero eigenvalues and a countably infinite number will be zero. This is the case with most integral equations—that the eigenvalues have some spectrum. This spectrum is discrete and countable for an integral equation defined on a finite interval. Thus, there are only a finite number of non-zero eigenvalues for the Karhunen-Loeve expansion. This is to say that only a finite number of temporal modes carry energy. In the limit of a very coherent process, there will be only one very long mode carrying energy. In the thermal source limit, we will show in section 7.5 that the temporal modes are approximately rectangles of length $\tau_c$ (where $\tau_c$ is the order of the optical period), all of which carry about equal energy (Van Trees 1968). In this thermal source limit as well, the noncorrelation of the $a_i$ cause them to become independent random Gaussian variables.

The Karhunen-Loeve equations can be transformed to an ordinary differential equation for an important class of functions when the correlation function admits stationary increments, allowing us to use the Wiener-Khintchine theorem (Davenport and Root 1958). We’ll presently show that we can convert (3.180) to a differential eigenvalue equation whenever we can express (or accurately approximate) the function $S_v(\omega)$ in the form

$$S_v(\omega) = \frac{N((i\omega)^2)}{D((i\omega)^2)},$$

(3.196)

where $N(x)$ is a polynomial of degree $n$, $D(x)$ is a polynomial of degree $d$, and $d > n$. That is to say that $S_v(\omega)$ is a rational function. Substituting (3.179) into the Karhunen-Loeve equation, we find that

$$\gamma_j \varphi_j(t_2) = \frac{1}{2\pi i} \int dt_2 \varphi_j(t_2) \int ds \frac{N(s^2)}{D(s^2)} e^{s(t_1-t_2)}. \tag{3.197}$$

We note that the effect of the $\partial/\partial t_1$ operator on this equation is

$$\gamma_j \frac{\partial \varphi_j(t_1)}{\partial t_1} = \frac{1}{2\pi i} \int dt_2 \varphi_j(t_2) \int ds \frac{sN(s^2)}{D(s^2)} e^{s(t_1-t_2)}. \tag{3.198}$$

If we use $\partial/\partial t_1$ and $\partial/\partial t_2$ as operators, we can replace $(\partial/\partial t_1)^2$ for $s^2$ in the argument of $D$ and $(\partial/\partial t_2)^2$ in the argument of $N$ to obtain

$$\gamma_j D(\frac{\partial^2}{\partial t_1^2}) \varphi_j(t_1) = \frac{1}{2\pi i} \int dt_2 \varphi_j(t_2) \int ds N\left(\frac{\partial^2}{\partial t_2^2}\right) e^{s(t_1-t_2)}. \tag{3.199}$$

Noting that

$$\int_{-\infty}^{\infty} e^{s(t_1-t_2)} ds = 2\pi \delta(t_1 - t_2), \tag{3.200}$$

we note that we can write

$$\frac{1}{2\pi i} \int dt_2 \varphi_j(t_2) \int ds N(s^2)e^{s(t_1-t_2)} = \int dt_2 \varphi_j(t_2) N\left(\frac{\partial^2}{\partial t_2^2}\right) \delta(t_1 - t_2), \tag{3.201}$$

which, from the definition of derivatives of the delta function, can be written as

$$\int dt_2 \varphi_j(t_2) N\left(\frac{\partial^2}{\partial t_2^2}\right) \delta(t_1 - t_2) = N\left(\frac{\partial^2}{\partial t_1^2}\right) \varphi_j(t_1), \tag{3.202}$$

which yields our resulting differential equation,

$$\gamma_j D\left(\frac{\partial^2}{\partial t_1^2}\right) \varphi_j(t) = N\left(\frac{\partial^2}{\partial t_1^2}\right) \varphi_j(t), \tag{3.203}$$
which is an ordinary differential equation of order 2 which will only have solutions for certain values of the parameter $\gamma_j$. That is, it is an eigenvalue equation.

It can be of interest at this point to consider an example, one for which we will borrow heavily from the book of Van Trees (1968). Let’s take for the spectral density of our process the form
\[ S(\omega) = 2\Delta \omega^2 P \omega^2 + (\Delta \omega)^2, \] (3.204)
which can be transformed (i.e. via the Wiener-Khintchine theorem) to yield
\[ R(\tau) = Pe^{-\Delta \omega |\tau|}. \] (3.205)

This Lorentzian lineshape is a good model for the linewidth of a laser in general or for the linewidth of a single mode of a thermal source or light-emitting diode (LED). The integral equation of interest then becomes
\[
\int_{-\tau_d/2}^{\tau_d/2} Pe^{-\Delta \omega |t_1-t_2|} \varphi(t_2) dt_2 = \lambda \varphi(t_1),
\] (3.206)
where the symmetric limits have been taken to simplify the calculation but really without any significant loss of generality. To remove the absolute value sign, we can rewrite the integral in the form
\[
\lambda \varphi(t_1) = \int_{-\tau_d/2}^{\tau_d/2} Pe^{-\Delta \omega |t_1-t_2|} \varphi(t_2) dt_2 + \int_{t_1}^{\tau_d/2} Pe^{\Delta \omega (t_1-t_2)} \varphi(t_2) dt_2.
\] (3.207)

A time derivative on $t_1$ of the above expression leads to
\[
\lambda \dot{\varphi}(t_1) = -\Delta \omega Pe^{-\Delta \omega t_1} \int_{-\tau_d/2}^{\tau_d/2} e^{\Delta \omega t_2} \varphi(t_2) dt_2 + \Delta \omega Pe^{\Delta \omega t_1} \int_{t_1}^{\tau_d/2} e^{-\Delta \omega t_2} dt_2.
\] (3.208)

A second $t_1$ derivative then yields
\[
\lambda \ddot{\varphi}(t_1) = \Delta \omega^2 P \int_{-\tau_d/2}^{\tau_d/2} e^{-\Delta \omega |t_1-t_2|} \varphi(t_2) dt_2 - 2\Delta \omega P \varphi(t_1).
\] (3.209)

But, substituting from the original equation, we obtain
\[
\lambda \ddot{\varphi}(t_1) = \Delta \omega^2 \lambda \varphi(t_1) - 2\Delta \omega P \varphi(t_1),
\] (3.210)
where the boundary conditions here on the function are quite complicated in that, to obtain them, we would need to substitute back into the integral equation, and the boundary conditions will be functions of the form of the solution. The optimum method then becomes seeing what solutions are possible to the differential equation independent of boundary conditions and then substituting that form back into the integral equation to see if it works (that is, making an ansatz). We see that the above differential equation will have sinusoidal solutions in the regime
\[
0 < \lambda < \frac{2P}{\Delta \omega}.
\] (3.211)

For $\lambda$ in this regime, we can then write that
\[
\varphi(t) = c_1 e^{jbt} + c_2 e^{-jbt},
\] (3.212)
where
\[ b^2 = -\Delta\omega^2(\lambda - 2P/\Delta\omega) \quad ; \quad 0 < b^2 < \infty. \] (3.213)

Plugging back into the integral equation, one finds
\[
\lambda c_1 e^{jbt_1} + \lambda c_2 e^{-jbt_1} = \int_{-\tau_d/2}^{\tau_d/2} P e^{-\Delta\omega t_1} e^{\Delta\omega t_2} (c_1 e^{jbt_2} + c_2 e^{-jbt_2}) dt_2
\]
\[ + \int_{t_1}^{t_2} P e^{\Delta\omega t_1} e^{-\Delta\omega t_2} (c_1 e^{jbt_2} + c_2 e^{-jbt_2}) dt_2, \] (3.214)
which can be simplified to
\[
\lambda c_1 e^{jbt_1} + \lambda c_2 e^{-jbt_1} = P \left\{ c_1(e^{jbt_1 - \Delta\omega t_1 e^{-j\tau_d/2} e^{-(j\tau_d/2)}})
\]
\[ + c_2(e^{jbt_1 - \Delta\omega t_1 e^{-j\tau_d/2} e^{-(j\tau_d/2)}}) \frac{\Delta\omega - jb}{\Delta\omega + jb}
\]
\[ + c_1(e^{\Delta\omega t_1 e^{-(j\tau_d/2) e^{(-j\tau_d/2)}}}) + c_2(e^{\Delta\omega t_1 e^{-(j\tau_d/2) e^{(-j\tau_d/2)}}}) \right\}, \] (3.215)
which, after some manipulation, can be rewritten as
\[
e^{-\Delta\omega t_1} \left[ \frac{c_1 e^{-(\Delta\omega + jb)\tau_d/2}}{\Delta\omega + jb} + \frac{c_2 e^{-(\Delta\omega - jb)\tau_d/2}}{\Delta\omega - jb} \right]
\[ = e^{\Delta\omega t} \left[ \frac{c_1 e^{-(\Delta\omega - jb)\tau_d/2}}{-\Delta\omega + jb} + \frac{c_2 e^{-(\Delta\omega + jb)\tau_d/2}}{-\Delta\omega - jb} \right] = 0. \] (3.216)

It is possible to show that we need
\[ c_1 = \pm c_2 \] (3.217)
where, for \( c_1 = -c_2 \),
\[ \tan \frac{\tau_d b}{2} = -b/\Delta\omega \] (3.218)
and, for \( c_1 = c_2 \),
\[ \tan \frac{\tau_d b}{2} = \Delta\omega/b \] (3.219)
with the corresponding eigenvalues
\[ \lambda_i = \frac{2P\Delta\omega}{(\Delta\omega^2 + b_i^2)^{1/2}} \] (3.220)
and where
\[ \varphi_i(t) = \begin{cases} \frac{\sqrt{2}}{\tau_d^{1/2} (1 + \sin \beta_i t/\tau_d)} \cos(b_i t/2), & i \text{ odd} \\ \frac{\sqrt{2}}{\tau_d^{1/2} (1 + \sin \beta_i t/\tau_d)} \sin(b_i t/2), & i \text{ even.} \end{cases} \] (3.221)

What happens when \( \lambda \) is outside the sinusoidal interval?, one could ask. In this regime, the solutions are hyperbolic. The easiest way to access this regime would be to perform the transformation that
\[ b = \mp ib'. \] (3.222)
If we make this substitution into the eigenvalue relation
\[ \tan \left( \frac{\tau_d b'}{2} \right) = \frac{\Delta\omega}{v_i}, \] (3.223)
we find that
\[ \tanh\left(\frac{\tau_d b_i'}{2}\right) = -\frac{\Delta \omega}{b_i}, \]  
(3.224)
which has no solutions. Likewise, if we make the substitution into the eigenvalue relation for the odd modes,
\[ \tan\left(\frac{\tau_d b_i'}{2}\right) = -\frac{b_i'}{\Delta \omega}, \]  
(3.225)
we will likewise find that
\[ \tanh\left(\frac{\tau_d b_i'}{2}\right) = -\frac{b_i'}{\Delta \omega}, \]  
(3.226)
which also has no solutions. Therefore, our only eigensolutions will have \( \lambda \) in the above-defined range. Perhaps more important than the shapes and specific values is the question of how many. For example, under what set of constants do we have but one \( \lambda \)?

Perusal of the two tangent equations above indicates that we can have a \( b_1 \) with value near zero. However, \( b_2 \) needs to minimally be \( \pi/2 \). We note from (3.198) that
\[ \sum_i \lambda_i = P_{\tau_d}. \]  
(3.227)
With this we could define a new normalized eigenvalue \( \mu_i \) by
\[ \mu_i = \frac{\lambda_i}{P_{\tau_d}} \]  
(3.228)
and therefore
\[ \sum_i \mu_i = 1. \]  
(3.229)
We could then rewrite the \( \lambda_i \) equation as
\[ \mu_i = \frac{2\Delta \omega \tau_d}{((\Delta \omega \tau_d)^2 + (b_i \tau_d)^2)^{\frac{3}{2}}}. \]  
(3.230)
We note that, from
\[ \tan\left(\frac{b_i \tau_d}{2}\right) = \frac{\Delta \omega}{b_i}, \]  
(3.231)
we see that
\[ (b_i \tau_d)^2 < 2\Delta \omega \tau_d, \]  
(3.232)
but in the limit of \( \Delta \omega \tau_d \) small, (3.232) will approach an equality for the first mode, and therefore in this limit (3.230) will reduce to
\[ \lim_{\Delta \omega \tau_d \to 0} \mu_1 \to 1. \]  
(3.233)
That is, for a narrow line source, the lowest-order eigenmode will carry all of the power. We will use the result again during our discussion of LED and laser linewidths in section 7.1.

Being able to find correlation times (which in effect determine the number of temporal modes) and being able to find spectral densities are useful, but ensembles are still bothersome. In a laboratory, we generally will have a single stochastic process going on—not an ensemble. Now we need to build a bridge from mathematics to practice. The ergodic theorem offers us a way out, when it holds, which it never does for information-carrying processes. Despite this fact, we can simplify descriptions under certain weaker assumptions like
local stationarity. (See section 11.1 of Chapter 11.) The idea is as illustrated in Figure 3.8. In general, the angular brackets we use to denote an average denote an average down an ensemble of realizations, such as the ensemble $v_1(t) \cdots v_n(t)$. Unfortunately, in practice we do not have an ensemble but only a realization which goes on for a long time (hopefully). For an ergodic process, one finds that the process of averaging down the ensemble should be identical to a “long” (read infinite) time average or, formally,

$$\bar{v} = \langle v(t) \rangle = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} v(t') \, dt'.$$

(3.234)

If we want to be practical, of course, we cannot go off to infinity, and this is illustrated also in Figure 3.8. Clearly there, one can consider $v(t)$ to be stationary between $0$ and $t_1$, between $t_1$ and $t_2$, between $t_2$ and $t_3$, etc. This is the case we will have in practice; that is, the transmitter in a digital system will be transmitting zeroes and ones. The picture itself clearly assumes that our detector has a response time fast compared to a bit period (probably not too much faster, though, as we don’t want our system to be too expensive). The bit period, quite in general, will greatly exceed the time between spontaneous emission events in the laser cavity, and therefore the average of a bit period should yield the same results as an average over the ensemble of event realizations that could occur given the laser bias level. When we have local stationarity, averages can be written as time averages of the form

$$\langle f(t) \rangle = \frac{1}{\tau} \int_{t-\tau}^{t} f(t') \, dt'$$

(3.235)

with $\tau$ judiciously chosen. We will use different forms of averaging essentially interchangeably through the development. For example, in section 11.1 we will use a different form of averaging to spectrally separate the signal from its signal-dependent shot noise.
3.5 Random Fields and Coherence Theory

Certainly, after a detector has integrated over some spatial area and turned the incident field into a current stream in a (approximately one-dimensional) wire, we really have a function of only time. However, the incident field also has a spatial dependence. One can consider such a field as a collection of time functions, one at each \((x, y, z)\) point in space, where at a single point the stochastic realization may appear as in Figure 3.9. If the scalar field is written as \(\psi(x, y, z, t)\), then at a specific \((x, y, z)\) one can write that

\[
\bar{\psi}(x, y, z) = \langle \psi(x, y, z, t) \rangle,
\]

(3.236)

where the averaging \(\tau\) of (3.235) is suitably chosen. One could perform the average at each point in space if one wanted to in order to obtain a time-independent average spatial field.

How might a field become random? As was discussed in the introduction to this chapter, quantum randomness comes right from the source. Spontaneous emission, the process which triggers laser action, is in itself random and leads to both high-frequency signal granularity and low-frequency phase diffusion. Multimode fiber propagation has a more direct effect on the spatial distribution. As modes all travel with different phase velocities, the spatial interference pattern between the modes varies in a complex manner as energy propagates down the fiber. (Path length differences can have an equivalent effect on the different modes involved in free-space propagation.) One could say that the signal distortion is actually deterministic, as one can calculate the propagation constants, but again reason must prevail. Typically, a multimode fiber (62.5 \(\mu\)m core, 0.26 numerical aperture) has 1000 modes (counting each polarization and each sine and cosine state). Further, a typical propagation distance may be a kilometer, which is a billion wavelengths. Further, the propagation constants are first-order-dependent on the fiber parameters, which are hard to specify to better than 1%, and further, these parameters can vary several percent as one moves longitudinally down the fiber. For example, to really predict a speckle pattern (modal interference pattern) a kilometer down a fiber would therefore require painstaking characterization of the fiber index profile every wavelength or so for a billion wavelengths. As the index measurements would be destructive, one could make a prediction of the pattern but could never check it, and all this assumes that the ambient conditions (temperature, pressure, humidity, etc.) are all kept totally constant under the process. (Atmospheric scintillation is the real spoiler in free-space propagation.) Good luck! The point is, again, that the speckle pattern contains too much information and therefore, by rule of reason, can be considered to be noise.

How do we generally treat fields in optics? We’ll see in much more detail in Chapters 5 and 9, but for the moment, consider a waveguide such as that depicted in Figure 3.10. We can describe propagation in such a guide by writing Maxwell’s equations (equations 2.39). If we make enough approximations, we can generally derive a scalar wave equation (see Chapters 2, 5, and 9)

\[
\nabla^2 \psi(x, y, z, t) + \frac{n^2(x, y; z)}{c^2} \frac{\partial^2 \psi(x, y, z, t)}{\partial t^2} = 0,
\]

(3.237)
where the $n^2(x, y; z)$ notation indicates that $n$ is a function of $x$ and $y$ but is only parameterized by $z$. That is, the variation with $z$ is so slow that we can ignore derivatives of it. Generally, the light will be narrowband, and further, as it is $z$-propagating down a nominally $z$-independent structure, we can write that

$$\psi(x, y, z, t) = \Re \left[ \psi(x, y)e^{i\beta z}e^{-i\omega t} \right]$$

(3.238)

to find that

$$\nabla^2 \psi(x, y, z) + (\beta^2 - k^2(x, y; z))\psi(x, y) = 0,$$

(3.239)

where the notation $(x, y; z)$ means that one can assume there is variation with $z$ but that it must be on a much slower scale than that of the transverse variation. If $k^2(x, y)$ is “nice,” the operator in (3.239) is “nice,” then $\psi$ will have an expansion in terms of orthogonal functions $\psi_i(x, y)$ where the $\psi_i(x, y)$ are normalized such that

$$\int \psi_i(x', y')\psi_j(x', y') \, dx' \, dy' = 2\eta \delta_{ij},$$

(3.240)

where $\eta_i$ is defined by

$$\eta_i = \sqrt{\mu_0 \epsilon_i} = \sqrt{\mu_0 \frac{1}{\epsilon_i}} = \eta_0 \frac{k_0}{\beta_i}.$$

(3.241)

With all of this, one can express $\psi_i(x, y, z, t)$ in the form

$$\psi(x, y, z, t) = \Re \left[ \sum_i a_i \psi_i(x, y)e^{i\beta_i z}e^{-i\omega t} \right],$$

(3.242)

where realistically the $\phi_i$’s and $\beta$’s will have only weak dependence on $z$. A problem is that there is nothing stochastic about (3.242) at all. It is just an orthogonal expansion, as we discussed in Chapter 2. How do we jam it in the stochasticness?

We have already made the assumption that the $\psi_i(x, y)$ form a complete set at any fixed $z$. Therefore, any disturbance at a fixed $z$ is expressible as (3.242) if one suitably chooses the $a_i$’s. The thing that we call stochastic comes from two sources, really: the spitting nature of the source and propagation. If we then let $a_i$ become a (stochastic) function of $z$ and $t$, we can take all we want into account by simply writing (see section 2.3 for some discussion)

$$\psi(x, y, z, t) = \Re \left[ \sum_i a_i(z, t) \psi_i(x, y)e^{i\beta_i z}e^{-i\omega t} \right].$$

(3.243)

A similar expansion to that made in (3.241) can also be made in free space. In free space, one knows that the eigenfunctions are just plane waves, and therefore one can write that

$$\psi(x, y, z, t) = \Re \left[ \int_0^\infty dk_x \int_0^\infty dk_y a(z, t, k_x, k_y)e^{i k_x x}e^{i k_y y}e^{\sqrt{k^2 - k_x^2 - k_y^2} z}e^{-i\omega t} \right],$$

(3.244)
where the modes for which \( k_x^2 + k_y^2 > k_0^2 \) are the so-called evanescent modes which are necessary to satisfy boundary conditions. (This type of expansion is discussed in much more detail in section 14.3.1. It should be noted that the expansion of (3.243) also must contain plane wave-like radiation modes as well as evanescent modes.) We have again stuffed all the perturbation and stochastic nature into the \( a(z, t, k_x, k_y) \). The problem with either of the forms (3.243) and (3.244) is that the \( a \)’s are not going to be independent random processes. We need to solve the Karhunen-Loeve equations for spatial modes as well as temporal modes to even know how many statistical degrees of freedom we have.

How do we characterize the spatial coherence of a random field? As is discussed in, for example, Mickelson (1992), coherence is generally operationally defined in terms of two experimental apparatus. Temporal coherence is defined by Michelson’s interferometer (see section 14.1 also) and spatial coherence by Young’s diffractometer. Formally, one would define the temporal correlation functions (for scalar fields) by

\[
R_\tau(\tau) = \langle \psi(t)\psi^*(t + \tau) \rangle. \tag{3.245}
\]

Assuming \( \psi \) to be locally time-stationary, one can write

\[
R_\tau(\tau) \approx \int_{t-\tau}^{t} \psi(t')\psi^*(t' + \tau) \, dt', \tag{3.246}
\]

which in general may appear as in Figure 3.11. There is generally a carrier of the form \( e^{-i\omega \tau} \) (assuming a narrowband signal) and some envelope, which is due to stochastic perturbation. Generally, the temporal coherence time is defined by the value of \( \tau \) at which the envelope has fallen to some fraction of its maximum (which must always lie at the origin from the Schwartz inequality) or, as we discussed earlier, we can make the concept more quantifiable if we can solve the Karhunen-Loeve equations for the temporal eigenfunctions.

Temporal coherence refers to a measurement that takes place at a given point on the wavefront but which requires correlating the field at one time with the field at the same point at a later time. Spatial coherence refers to a correlation measurement which correlates the field at two different spatial points at the same time. A formal expression for this correlation could be written in the form

\[
R_{\psi(r_1, \psi(r_2)}(r_1, r_2) = \langle \psi(r_1, t)\psi^*(r_2, t) \rangle, \tag{3.247}
\]

where the \( r_1 \) and \( r_2 \) now indicate the \( x \) and \( y \) coordinates transverse to \( z \). Assuming stationarity, one often writes

\[
\langle \psi(r_1, t)\psi^*(r_2, t) \rangle = \int_{t-\tau}^{t} \psi(r_1, t')\psi^*(r_2, t') \, dt'. \tag{3.248}
\]

In general, the expression in (3.246) will be a function of absolute as well as relative coordinate, but as it may take in some region (or regions) of the disturbance, the field appears to be statistically homogeneous. If this were the case, one could define

\[
d = |r_1 - r_2| \tag{3.249}
\]

and find that

\[
R_{\psi(r_1, \psi(r_2)}(d) = \int_{t-\tau}^{t} \psi(r_1, t')\psi^*(r_2, t') \, dt' \tag{3.250}
\]

to find that \( R \) may have a realization which appears much as that depicted in Figure 3.12. Figure 3.12 appears much as Figure 3.11 except with space replacing time and \( k \) replacing \( \omega \) in the expression for the carrier. Spatial coherence length would be defined analogously to temporal coherence time, as was discussed in relation to Figure 3.11. Indeed, as before, we could use the Karhunen-Loeve expansion to make the
concept more definable by having an equation defining spatial coherence eigenmodes. As we proceeded with temporal modes, we could proceed here with spatial modes to write that
\[ \psi(r_1, t_1) = \sum_i a_i \varphi_i(r_1, t_1), \]
(3.251)
where the independent condition requires that
\[ \langle a_i a_j^* \rangle = \delta_{ij} \gamma_j \]
(3.252)
and the orthogonality condition that
\[ \int \varphi_i(r, t) \varphi_j(r, t) d^3r dt = \delta_{ij}. \]
(3.253)
Writing that
\[ R(r_1, t_1, r_2, t_2) = \langle \psi(r_1, t_1) \psi^*(r_2, t_2) \rangle, \]
(3.254)
we see that we can write that
\[ \int R(r_1, t_1, r_2, t_2) \varphi_k(r_2, t_2) d^3r_2 dt_2 = \int \sum_{ij} \langle a_i a_j^* \rangle \varphi_i(r_1, t_2) \varphi_j^*(r_2, t_2) \varphi_k(r_2) dt_2. \]
(3.255)
Using independence and orthogonality in the above yields the result that
\[ \int R(r_1, t_1, r_2, t_2) \varphi_k(r_2, t_2) d^3r_2 dt_2 = \gamma_k \varphi_k(r_1, t_1). \]
(3.256)

If we can make the identification of \( z \) as the propagation direction (plane wave-like propagation) and the medium is not time-varying, as will be discussed in section 5.2, then the temporal coherence propagates without change while the transverse coherence evolves with propagation. This is to say that the temporal and spatial coherences are independent. Practically, we will be considering waves which have a propagation direction \( z \). Further, we will be interested in transverse spatial modes in a plane of constant \( z \). Assuming the propagation is plane wave-like then allows us to separate (3.256) into the two equations:
\[ \int R(r_1, r_2) \varphi_k(r_2) d^2r_2 dt_2 = \gamma_k \varphi_k(r_1) \]
\[ \int R(t_1, t_2) \varphi_k(t_2) dt_2 = \gamma_k \varphi_k(t_1). \]
(3.257)
The last question that one might ask concerning coherence might be, how is the coherence affected as one propagates a wave? Clearly, as mentioned above, if a medium is non time-varying, the temporal coherence is unaffected. The spatial coherence will be, however. In fact, an incoherent source of finite extent will have a coherence length given by
\[ \ell_c \approx \frac{\lambda}{2NA} \]
(3.258)
after propagating through an optical system of numerical aperture \( NA \), a result given by the so-called Van Cittert Zernike theorem (Mickelson, 1992), as we will also derive later in section 5.3 of Chapter 5 and discussed further in Chapter 9. In general, spatial coherence will increase with propagation distance, as the “rays” which are not coherent with those going in a given direction will not go in that direction anyway. Elegant (but complex) formulae can be found for propagating correlation functions such as those defined by (3.247) and (3.248) from the use of diffraction theory as is done in Born and Wolf (1975), but the derivation of such formulae would take us too far afield here. We will perform a bare-bones analysis, though, of coherence propagation in section 5.2 of Chapter 5.
Figure 3.11: A possible temporal correlation function.

Figure 3.12: A sketch of what a spatial correlation function may look like.
Basically, with reference to our general systems model of the first chapter, our picture is that our information stream is to somehow be used to gate on and off a stochastic process which is the optical source. This gated stochastic process is then propagated through the system to the detector as a random field which is continuously transformed by the optical system until arrival at the detector. The detector will then convert the incident random field to a current, which is again a stochastic process. In a digital system, an integrator with a time constant of the bit time will then turn this process into a random number which is randomly distributed according to a distribution which is a function of all the elements of the communications system. It is the properties of this distribution that will determine the salient features of a digital system as a whole. In an analog system, our main measure of fidelity is the average signal-to-noise ratio, which becomes a strong function of the decoding circuitry.

3.6 Summary of the Probability Densities Most Commonly Used in Optical Communications

The most commonly used continuous densities in optical communications are those illustrated in Figures 3.3 and 3.4—that is, the Gaussian, the uniform, the delta, the log normal, the Rayleigh, the Rician, the chi-squared, the gamma, and the beta.

The Gaussian density is expressible as

$$p_x(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x - \langle x \rangle}{\sigma}\right)^2}$$

and has a characteristic function of the form

$$\psi_x(\omega) = e^{j\omega\langle x \rangle} e^{-\frac{\omega^2}{2}}$$

and is important as it is the density of any variable $x$ which is the result of a “complicated enough” process. That is, there is a central limit theorem which was discussed in the last section, which states that, if an $x$ is the sum of enough individual $x_i$’s, then its distribution is Gaussian.

The uniform density is expressible as

$$p_x(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

and has a characteristic function of the form

$$\psi_x(\omega) = e^{j\omega\langle x \rangle} \sin \frac{\omega(b-a)}{2}$$

and is indicative of a white noise-type process. That is, within the regime of interest ($a < x < b$), no value is any more likely than any other value.

The delta density is expressible as

$$p_x(x) = \delta(x - a)$$

and has a characteristic function of the form

$$\psi_x(\omega) = 1$$

and is the density of a deterministic variable $x$ that is one that has a fixed value. In cases where there is a density which is a sum of delta functions, we say that we have a discrete random variable if the coefficients of the various delta functions are random variables. We oftentimes call the set of coefficients the probability distribution of the random variable.
The log normal density in the variable $y$ is obtained by making the transformation $y = \exp\{x\}$ in the Gaussian probability distribution to obtain

$$p_x(x) = \frac{1}{\sqrt{2\pi} \sigma x} \exp\left\{ \frac{\ln x + \sigma^2}{2\sigma^2} \right\}^2.$$ \hspace{1cm} (3.265)

The characteristic function for this density is not obtainable in closed form. This density is used to describe the amplitude of a free-space wave propagating through a scintillating atmosphere or the amplitude of a wave received from multiple interfering paths.

The Rayleigh density,

$$p_x(x) = \frac{x}{\sigma^2} e^{\frac{x^2}{2\sigma^2}},$$ \hspace{1cm} (3.266)

is not Fourier transformable to a closed form. However, a distribution derived from it, the exponential distribution (see equation (3.278)), is. The Rayleigh density can be used to describe the amplitude statistics of a low coherence (thermal or LED-type) source.

The Rician density,

$$p_x(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} I_0\left(\frac{x(x^2)^{1/2}}{\sigma^2}\right),$$ \hspace{1cm} (3.267)

is, as with the Rayleigh density, not Fourier transformable in a closed form, so that the characteristic function is not an elementary function. However, a distribution derived from it (see equation (3.280)) is. The Rician distribution describes the amplitude statistics of a source of amplitude $x$, but corrupted with additive Gaussian noise of standard deviation $\sigma^2$. This distribution could represent a stream of laser light copropagated with starlight or blackbody radiation.

The $t$ distribution,

$$p_x(x) = \frac{\Gamma\left(\frac{M+1}{2}\right)}{\sqrt{M\pi} \Gamma\left(\frac{M}{2}\right)} \left(\frac{x^2}{M+1}\right)^{(M+1)/2},$$ \hspace{1cm} (3.268)

reduces to the Gaussian distribution in the limit of large $M$. It is the distribution we would expect to obtain if one measured $M$ samples from a Gaussian (normal) distribution. This distribution is important in the theorem of sampling as well as the design of experiments. The characteristic function is not easily obtainable.

The gamma density,

$$p_x(x) = \begin{cases} \frac{x^{b-1} e^{-cx}}{\Gamma(b+1)}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$ \hspace{1cm} (3.269)

with a characteristic density function given by

$$\psi_x(\omega) = \frac{\omega^{b+1}}{(c - j\omega)^{b+1}},$$ \hspace{1cm} (3.270)

is one that can be used to describe integrated intensity statistics of a portion of a speckle pattern which is small compared to the total speckle pattern size. The beta density,

$$p_x(x) = \begin{cases} \frac{\Gamma(b+c+2)}{\Gamma(b+1)\Gamma(c+1)} x^b (1-x)^c, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$ \hspace{1cm} (3.271)

which is not Fourier transformable in closed form, is the generalization of the gamma distribution to the case where integrals used to find the integrated intensity statistics in the gamma distribution can be extended to cover the whole speckle distribution. In this sense, the beta density can be used to describe the statistics of the field received by a detector subtending most of the surface of a multimode optical fiber (Mickelson and
Weierholt, 1983; Hjelme and Mickelson, 1983). A discussion of modal noise in section 9.4 of Chapter 9 will employ these distributions.

The most commonly used discrete distributions will be the Poisson distribution and distributions derived from it through the conditional Poisson counting process. The Poisson distribution is given by

\[ p_k(k) = \frac{e^{-m} m^k}{k!} \]  

(3.272)

with a characteristic function given by

\[ \psi_k(\omega) = e^{m(e^{j\omega} - 1)} \]  

(3.273)

and the conditional Poisson by

\[ p_k(k) = \int pos(k, m) p_m(m) \, dm, \]  

(3.274)

with a characteristic function given by

\[ \psi_k(\omega) = \psi_m(-j(e^{j\omega} - 1)), \]  

(3.275)

where the \( p_m(m) \) is a count density characterizing some random conditioning process which will modify the pure spontaneous emission process that the Poisson process represents in the limit where the detector is “fast.” When \( p_m(m) \) is deterministic (delta distribution) with count parameter \( \overline{m} \), the conditional Poisson becomes an actual Poisson of parameter \( \overline{m} \), and this distribution represents the counts from a good laser source, a distribution which represents standard shot noise.

The Bose-Einstein count distribution is expressible as

\[ p_k(k) = \frac{1}{1 + \overline{m}} \left( \frac{\overline{m}}{1 + \overline{m}} \right)^k \]  

(3.276)

with a characteristic function given by

\[ \psi_k(\omega) = \frac{1}{1 + \overline{m}(1 - e^{j\omega})}, \]  

(3.277)

where the \( \overline{m} \) is the average value of the count parameter in the count distribution

\[ p_m(m) = \frac{1}{\overline{m}} e^{-m/\overline{m}}, \]  

(3.278)

which can be derived from the Rayleigh distribution by a transformation of the form

\[ m = \frac{\eta}{h\omega} A_d \tau_d x^2, \]  

(3.279)

where \( \eta \) is the quantum efficiency, \( h\omega \) the photon energy, \( A_d \) the detector area, and \( \tau_d \) the detector time, which is the type of count distribution emitted by a thermal source if detected by an ultrafast detector.

When the Poisson distribution is conditioned by a count distribution \( p_m(m) \) obtained from the Rician density, which is given by

\[ p_m(m) = \frac{1}{\overline{m}_i} \exp \left\{ -\frac{m + \overline{m}_i}{\overline{m}_i} \right\} I_0 \left( \frac{(m\overline{m}_i)^{1/2}}{\overline{m}_i} \right), \]  

(3.280)

with characteristic function

\[ \psi_m(\omega) = \frac{1}{1 - j\omega \overline{m}_i} \exp \left\{ \frac{j\omega \overline{m}_i}{1 - j\omega \overline{m}_i} \right\}, \]  

(3.281)

one obtains the Laguerre counting probability

\[ p_k(k) = \frac{1}{1 + \overline{m}_i} \left( \frac{\overline{m}_i}{1 + \overline{m}_i} \right)^k \exp \left\{ -\frac{\overline{m}_i}{1 + \overline{m}_i} \right\} L_k \left( \frac{-\overline{m}_i}{1 + \overline{m}_i} \right), \]  

(3.282)
where the characteristic function is given by
\[
\psi_m(\omega) = \frac{1}{1 + \frac{m}{\mu_i}(1 - e^{i\omega})} \exp \left\{ \frac{\mu_c(e^{i\omega} - 1)}{1 + \frac{m}{\mu_i}(1 - e^{i\omega})} \right\},
\] (3.283)
where the Laguerre polynomial is defined as
\[
L_k(x) = \sum_{j=0}^{k} \binom{k}{j} (-x)^j j!.
\] (3.284)

Oftentimes one defines the associated Laguerre probability function,
\[
\text{Lag}[k, \mu_c, \mu_i, c] = \frac{m^k}{(1 + \frac{m}{\mu_i})^{k+c+1}} \exp \left\{ -\frac{\mu_c}{1 + \frac{m}{\mu_i}} \right\} L_k \left[ \frac{-\mu_c}{\mu_c(1 + \frac{m}{\mu_i})} \right],
\] (3.285)
where \(L_k(x)\) is the generalized Laguerre polynomial of index \(k\) and order \(c\), given by
\[
L_k(x) = \sum_{i=0}^{k} \binom{k+c}{i} (-x)^i i!.
\] (3.286)

A generalization of the Bose-Einstein count statistics leads to the negative binomial variate density. We should recall that temporal eigenmodes of a stochastic signal can be defined from the Karhunen-Loeve (K-L) expansion of (3.180). A temporal mode will in essence have a width equal to its coherence time \(\tau_c\), which we will show from the Karhunen-Loeve expansion in section 7.5 using the correlation function
\[
R(\tau) = R_0 \sin \frac{\tau}{\tau_c}.
\] (3.288)

Although the Karhunen-Loeve equations are exactly solvable for the coherence function above, as we will discuss in section 7.5, a useful approximation is to say that each temporal mode has a flat amplitude for \(\tau_c\) and is zero elsewhere. If we then say that the receiver has a bandwidth such that it is matched to a rectangular pulse of width \(\tau_r\), then the receiver will receive \(\frac{\tau_r}{\tau_c} + 1\) independent modes during its period \(\tau_r\). These temporal modes will sum in the receiver so that the received signal will be a sum of the \(N = \frac{\tau_r}{\tau_c} + 1\) independent modes. As each mode has a \(\psi_k(\omega)\) of the form of (3.272), the composite \(\psi_k(\omega)\) will be
\[
\psi_k(\omega) = \left[ \frac{1}{1 + \frac{m}{\mu_i}(1 - e^{i\omega})} \right]^N.
\] (3.288)

The inverse transform of this expression, which can be obtained either by complex plane integration or by an \(N\)-fold convolution of the Bose-Einstein distribution, yields the negative binomial variate distribution
\[
p_k(k) = \binom{N+k-1}{k} \left( \frac{1}{1+\frac{m}{\mu_i}} \right)^N \left( \frac{\mu_i}{1+\frac{m}{\mu_i}} \right)^k.
\] (3.289)

This is the count distribution which would correspond to that received from a reasonably incoherent source such as an LED. It is interesting to note that, in the limit where \(N\) becomes large but \(Nm\) is fixed such that \(m\) becomes small, the negative binomial variate distribution becomes Poisson.

A generalization of the Laguerre counting probability is the one obtained when the received signal consists of a deterministic signal corrupted by \(N\) temporal modes of Gaussian noise. This distribution is generally referred to as the \(N\)-times convolved Laguerre counting probability density. In this, the conditioning distribution \(p_m(m)\) is the noncentral chi-squared density of degree \(N\),
\[
p_m(m) = \frac{1}{\mu_i} \left( \frac{m}{\mu_c} \right)^{N-1} \exp \left\{ -\frac{m + \mu_c}{\mu_i} \right\} I_{N-1} \left( \left( \frac{m \mu_c}{\mu_i} \right)^{1/2} \right).
\] (3.290)
with characteristic function

\[ \psi_m(\omega) = \left[ \frac{1}{1 - m_i j\omega} \exp \left\{ \frac{m_c j\omega}{1 - m_i j\omega} \right\} \right]^N, \]  
(3.291)

which yields a count distribution

\[ p_k(k) = \frac{m^k}{(1 + m_i)^{k+N}} \exp \left\{ \frac{m_c}{1 + m_i} \right\} L_{k-1}^{N-1} \left[ \frac{m_c}{1 + m_i} \right] \]  
(3.292)

with characteristic function

\[ \psi_m(\omega) = \left[ \left( \frac{1}{1 - \bar{m}_i (e^{j\omega} - 1)} \right) \exp \left\{ \frac{\bar{m}_c (e^{j\omega} - 1)}{1 - \bar{m}_i (e^{j\omega} - 1)} \right\} \right]^N. \]  
(3.293)

Before leaving the topic, it seems that some motivation should be given as to why we learn about all of these distributions. As one will note in the telecommunications literature, no matter what one calculates at the end, a usual approach is to invoke the law of large numbers, approximate the distribution as Gaussian, and use the Gaussian to calculate the bit error rate (BER), etc. A problem is that the law of large numbers applies to the central peak of the Gaussian, not the wings. The central peak corresponds to two or three standard deviations (\( \sigma \)'s). The confidence level at 3\( \sigma \) is roughly 99%, with a corresponding BER of roughly 1%. In telecommunications, one needs BERs of less than roughly 10^{-9}. This in itself is not in the central peak of the Gaussian. Still worse, however, at a 1-gigabit/second data rate, a BER of 10^{-9} corresponds to one error per second, which is only acceptable in a high-latency system employing a correction code. In data communications or in a distributed computation system, BERs of 10^{-15} or even 10^{-18} are necessary. The Gaussian approximation probably cannot be of much value there. More will be discussed of this, though, as the presentation continues; in particular, this topic will be taken up again in detail in section 12.2.

**Problems**

1. Show that

\[ E\{(x - a)^2\} = \sigma_x^2 + (a - m_1)^2 \]

for any \( a \) and that

\[ E\{(x - a)^2\} \]

is minimum if \( a = m_1 \).

2. Find the characteristic function \( \psi_x(\omega) \) for the Gaussian distribution

\[ P_x(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left[ - \frac{(x-m_1)^2}{2\sigma^2} \right]. \]

3. For the Poisson distribution

\[ P_k(k) = \frac{m^k}{k!} e^{-m} \]

where

\[ m = \int_0^T n(t') dt', \]

find the expectation \( E[k^3] \).

4. Recall that, for the Bose-Einstein distribution,

\[ p_k(k) = \frac{1}{1 + \bar{n}} \left( \frac{\bar{n}}{1 + \bar{n}} \right)^k \]
and
\[ \psi_k(\omega) = \frac{1}{1 + \bar{n}(1 - e^{i\omega})}, \]
where \( \bar{n} \) is the count parameter. Find
\( (a) \ E_k(k); \)
\( (b) \ E_k(k^2); \)
\( (c) \ \text{var}(k) \)
for this distribution. Recall that
\[ E_k(k^n) = \sum k^n P_k(k) \]
and
\[ \psi_k(\omega) = \sum m_n \frac{(j\omega)^n}{n!} \]
\[ m_n = E_k(k^n). \]

5. The Bose-Einstein count statistics are given by
\[ p_k(k) = \frac{1}{1 + m} \left( \frac{m}{1 + m} \right)^k \]
with a characteristic function
\[ \psi_m(\omega) = \frac{1}{1 + m(1 - e^{i\omega})}. \]
This distribution is not too useful for optical communications, however, as the coherence time of most purely spontaneous emission sources is so small. If we were to write that
\[ n = \frac{\tau_d}{\tau_c} + 1, \]
where \( \tau_d \) is the detector time and \( \tau_c \) the source coherence time, we would obtain the characteristic function
\[ \psi_m(\omega) = \left[ \frac{1}{1 + m(1 - e^{i\omega})} \right]^n. \]
There are at least two ways to invert this characteristic function. One is to note that the function \( f(\omega), \)
\[ f(\omega) = \frac{1}{1 + m(1 - e^{i\omega})}, \]
has a simple pole at
\[ \omega = j \ln \left( \frac{m}{m + 1} \right), \]
and therefore, from Cauchy’s residue formula for the inverse transform,
\[ p_k(k) = 2\pi i \lim_{\omega \to j \ln \left( \frac{m}{m + 1} \right)} \frac{1}{2\pi} \left( \omega - j \ln \frac{m}{m + 1} \right)^n \frac{\partial^n}{\partial \omega^n} \left[ \frac{1}{2\pi} \frac{e^{j\omega k}}{1 + m(1 - e^{j\omega})} \right]. \]
The other way would be to note that
\[ p_k(k) \bigg|_n = p_k(k) \bigg|_{BE} * p_k(k) \bigg|_{BE} * \cdots * p_k(k) \bigg|_{BE}, \]
where one would need to perform \( n \) discrete convolutions. These convolutions turn out to be quite simple and immediately yield the form
\[ p_k(k) = \left( \frac{1}{1 + m} \right)^n \left( \frac{m}{1 + m} \right)^k f(n, k), \]
but the \( f(n, k) \) requires some work and probably needs to be solved for recursively.
(a) Evaluate the \( n \)-th order pole residue to find the negative binomial variate distribution
\[
p_k(k) = \binom{n - 1 + k}{k} \left( \frac{m}{1 + m} \right)^k \left( \frac{1}{1 + m} \right)^n.
\]

(b) Perform the \( n \) discrete convolutions to set up a recursion for \( f(n, k) \) and then show
\[
f(n, k) = \binom{n - 1 + k}{k} = \frac{(n - 1 + k)!}{(n - 1)! k!}.
\]

6. The Laguerre counting probability density is given by
\[
p_k(k) = \frac{1}{1 + m_i} \left[ \frac{m_i}{1 + m_i} \right]^k \exp \left\{ -\frac{m_c}{1 + m_i} \right\} L_k \left( -\frac{m_c}{m_i(1 + m_i)} \right),
\]
where the Laguerre polynomial is given by
\[
L_k(x) = \sum_{j=0}^{k} \frac{k!}{j!} \frac{(-x)^j}{j!}
\]
and \( p_k(k) \) has an associated characteristic function
\[
\psi_k(\omega) = \frac{1}{1 + m_i(1 - e^{i\omega})} \exp \left\{ -\frac{m_c(1 - e^{i\omega})}{1 + m_i(1 - e^{i\omega})} \right\}.
\]

Find
(a) \( E(k) \)
(b) \( E(k^2) \)
(c) \( \sigma^2 \)
where
\[
E(f(k)) = \sum_{k=0}^{\infty} f(k) p_k(k)
\]
and where
\[
\psi_k(\omega) = \sum_{n=0}^{\infty} E(k^n) \left( \frac{j\omega}{n!} \right)^n.
\]
You might want to recall that
\[
\frac{1}{1 - x} = \sum_{n=0}^{\infty} x^n.
\]

7. The random variable \( N \) is Poisson distributed with parameter \( \alpha \) and random variable \( x \) independent of \( N \). Show that, if \( y = Nx \) and
\[
p_x(x) = \frac{\alpha}{\pi(a^2 + x^2)}
\]
then
\[
\phi_y(w) = \exp \{ \alpha e^{-\alpha|w|} - \alpha \}.
\]

8. Consider a random variable \( x \) defined in terms of a random variable \( t \) by
\[
x = \sin \omega t,
\]
where the probability density of $t$ is given by

$$p_t(t) = \begin{cases} \frac{\omega}{2\pi}, & -\frac{\pi}{\omega} < t < \frac{\pi}{\omega} \\ 0, & t > \left|\frac{\pi}{\omega}\right| \end{cases}$$

Find $p_x(x)$. (Hint:)

$$p_z(z) = \left. \frac{p_x(x)}{\partial f/\partial x} \right|_{x = f^{-1}(z)}$$

9. The light intensity in an optical beam is Gaussian. That is, the intensity $I(\sigma)$ is given by

$$I(r) = \frac{1}{\sqrt{2\pi} \sigma} e^{-r^2/2\sigma^2}.$$ 

Find the density function of the light amplitude distribution assuming that we sample the amplitude without knowing the position—that is, we uniformly sample the coordinate. The Gaussian mode is the fundamental mode, and therefore the amplitude can be considered as the square root of the Gaussian; phase is not important for a fundamental mode. (Hint:)

$$p_z(z) = \left. \frac{p_x(x)}{\partial f/\partial x} \right|_{x = f^{-1}(z)}$$

10. It is said in (Yadlowsky 1992) that Sudarshan has proven that the classical and quantum statistics of light do not differ to second-order statistics and that this has been proven by Hanbury-Brown and Twiss. (See (Mickelson 1992).) In terms of what we are doing in this course, classical light probably refers to the continuous distribution

$$p_k(k) = \frac{1}{\sqrt{2\pi}m} e^{-\frac{(k-m)^2}{2m}}$$

with characteristic function

$$\psi_k(\omega) = e^{im\omega} e^{-m\frac{\omega^2}{2}},$$

while quantum probably refers to light distributed according to the discrete distribution

$$p_k(k) = \frac{m^k}{k!} e^{-m}$$

with characteristic function

$$\psi_k(\omega) = e^{m(e^{i\omega} - 1)}.$$ 

Use the characteristic functions to find all of the statistics (the $m_n$’s as well as the central moments such as $\sigma^2 = E[(k - E(k))^2]$) up through fourth order and prove or disprove Sudarshan, Hanbury-Brown, Twiss, Yadlowsky, and me.
11. In finding the Gaussian approximation to the Poisson distribution, several times the limits of \( m \) being large and \( k \) being close to \( m \) are used.

(a) Estimate how badly in error (i.e. find a function giving the relative error in terms of \( m \) and \( m - k \)) the Gaussian approximation becomes as one moves out \( N\sigma \)'s from the mean, where \( \sigma = \sqrt{m} \).

(b) It is stated that, in the proof of the law of large numbers, the approximations are also made that there is a large number of samples and that the value of the distribution is close to its mean value. Where? Can you find a function giving you the error in the law of large numbers as a function of distance from the mean, as well as the total number of samples \( N \) in the summation?

12. Consider the Karhunen-Loeve equation

\[
\int_{-\tau_a/2}^{\tau_a/2} R(t_1, t_2) \varphi(t_2) \, dt_2 = \lambda \varphi(t_1). 
\]

(a) Let’s say that \( R \) is expressible in the form

\[
R(t_1, t_2) = p_0 f(t_1)f(t_2). 
\]

Solve for the \( \varphi \) and \( \lambda \).

(b) Let’s say

\[
R(t_1, t_2) = \sum_{n=0}^{N} p_i f_n(t_1)f_n(t_2) 
\]

where

\[
\int_{-\tau_a/2}^{\tau_a/2} f_n(t)f_m(t) \, dt = \delta_{nm}. 
\]

Solve for all the possible \( \varphi \)'s and \( \lambda \)'s of the above Karhunen-Loeve equation.

13. Recall the Karhunen-Loeve (K-L) equation, expressible as

\[
\int_{t-\tau_a}^{t} R(t_1, t_2) \varphi(t_2) \, dt_2 = \lambda \varphi(t_1). 
\]

Let’s say we don’t know how to solve it or, at least, the analytical solution is intractable. Let’s say we have a set of functions \( \psi_n(t) \) that are tractable and easy to numerically evaluate.

(a) Transform the above integral equation to a matrix eigenvalue problem for the linear combinations of the \( \psi_n(t) \) which approximate the exact eigenvalues.

The spatial K-L equations take the form

\[
\int_{\mathcal{A}} \int R(x_1, y_1, x_2, y_2) \varphi(x_2, y_2) \, dx_2 \, dy_2 = \lambda \varphi(x_1, y_1). 
\]

Let’s say we again have a complete set of approximate, easily handleable eigenfunctions \( \psi^*_n(x, y) \).

(b) Transform the spatial K-L equation to a matrix eigenvalue problem.

(c) Let’s say that the coherence function does not separate such that the K-L equation takes the form

\[
\int_{\mathcal{A}} \int_{t-\tau_a}^{t} R(t_1, x_1, y_1, t_2, x_2, y_2) \varphi(t_2, x_2, y_2) \, dt_2 \, dx_2 \, dy_2 = \lambda \varphi(t_1, x_1, y_1). 
\]
(d) How might one use the complete sets $\psi_m(t)$ and $\psi^n_{m}(x, y)$ to transform the K-L equation to a matrix eigenvalue problem?

14. A bandlimited white noise process can be defined by its spectral density

$$S(\omega) = \frac{\pi P}{\Delta \omega} \text{rect} \left( \frac{\omega}{\Delta \omega} \right)$$

which can be related by the Wiener-Khintchine theorem to its correlation function $R(t - t')$ to give

$$R(t - t') = P \frac{\sin \Delta \omega(t - t')}{\Delta \omega(t - t')}.$$  

The Karhunen-Loeve (K-L) equation for this process would then be

$$\lambda \varphi(t) = \int_{-\tau_d/2}^{\tau_d/2} P \frac{\sin \Delta \omega(t - t')}{\Delta \omega(t - t')} \varphi(t') dt'.$$

We could use normalized variables

$$c = \frac{\Delta \omega \tau_d}{2},$$

$$x = \frac{2t}{\tau_d},$$

$$\mu = \frac{\lambda}{P \tau_d}$$

to rewrite the K-L equation in the form

$$\mu f(x) = \int_{-1}^{1} \frac{\sin c(x - x')}{c(x - x')} f(x') dx'.$$

It is said in various places that successive differentation of the equation should give rise to the equation

$$(1 - x^2) \frac{d^2 f}{dx^2} - 2x \frac{df}{dx} + (\mu - c^2 x^2) f(x) = 0; \quad -1 < x < 1,$$

which defines the angular prolate spheroidal wave function $S_{0n}(c, x)$ with normalized eigenvalues

$$\mu_n = \frac{\lambda_N}{P \tau_d}$$

given by the radial prolate spheroidal wave function

$$\mu_n = 2 \left[ R^{(1)}_{0n}(c, 1) \right]^2.$$

(a) Can you derive the above differential equation from the integral equation?

(b) If the answer to (a) is yes, then carry out the derivation using enough steps that even I can follow your work.

(c) If the answer to (a) is no, carry out a derivation to show what you do get.

15. Say that the auto-correlation of a field disturbance is measured in a Michelson interferometer to be given by

$$\langle E(t + \tau)E(t) \rangle = I_0 e^{-\gamma |\tau|} e^{-i\omega \tau}.$$  

(a) How might one define the coherence time for this field?
(b) Find the expression for the spectral density \( S_\omega(\omega) \) for this field.

(c) How might one express the count statistics \( P_k(k) \) in the limits when the detector time \( \tau_d \) is

\[ \tau_d << \frac{1}{\gamma} \quad \tau_d >> \frac{1}{\gamma} \]

16. Consider a transition between two different fiber types at coordinate \( z = 0 \), as depicted below.

17. Say that, in fiber 1, the propagating field can be expanded into the form

\[ \psi_1(x, y, z) = \sum_n a_n^{(1)} \phi_n^{(1)}(x, y) e^{i\beta_n^{(1)}z} \]

and that, in fiber 2, the propagating field can be expanded into the form

\[ \psi_2(x, y, z) = \sum_m a_m^{(2)} \phi_m^{(2)}(x, y) e^{i\beta_m^{(2)}z} \]

where the \( a_n^{(i)} \)s are determined from excitation conditions, the \( \phi_n^{(i)} \)s are orthonormal functions in each fiber, and the \( \beta \)s are propagation constants. The time dependence has been suppressed but can be assumed separable and therefore the same in each spatial mode.

(a) Find the \( a_m^{(2)} \)s in terms of the \( a_n^{(1)} \)s and the eigenfunctions.

(b) What effect, if any, would multiple temporal modes have on the result in (a)?

(c) What effect, if any, would multiple spatial modes (K-L modes) have on the result in (a)?

(d) What effect, if any, would multiple spatial and temporal (K-L) modes have on the result in (a)? What effect would dispersion have on this result if the spatial (K-L) modes started to be separated by distances comparable to or greater than the source coherence time?

(c) (Hint: Use separation of variables to obtain an eigenfunction problem.)
Bibliography


