3

Theory of Random Processes

3.1 BASIC CONCEPTS

In this section we present a brief review of the basic definitions and concepts of probability theory. The reader who has no background in this area is advised to consult one of the many general texts on the subject, such as Davenport and Root (1958), Davenport (1970), or Papoulis (1965).

We make no claim of rigor or completeness in this review. Our goal is merely to assemble the tools necessary in an analysis of noise in radiographic systems.

3.1.1 Probability and Probability Distributions

In the most elementary terms, the probability of some event is its relative frequency of occurrence in a large number of trials. For example, if we cast a fair die 6,000,000 times, we should expect the side with three spots to show up about 1,000,000 times; the probability of rolling a three is $\frac{1}{6}$. More formally, if event A is a possible outcome of a certain experiment, and this event occurs $m(A)$ times during M repetitions of the experiment, then we may define the probability of occurrence of A as

$$\Pr(A) \equiv \lim_{M \to \infty} [m(A)/M]. \tag{3.1}$$

We denote probability as Pr( ) rather than the more common P( ) in order to avoid confusion with the point spread function $p(\mathbf{r})$ and its transform

$P(\boldsymbol{\rho})$, which were introduced earlier. Note also that Pr( ) does not imply a specific functional dependence, but is to be read "the probability that ... occurs."

It is evident from this definition that probability is a nonnegative real number. Furthermore, if an event is certain to occur in every repetition of the experiment, the definition shows that its probability is one. Therefore,

$$0 \leqslant \Pr(A) \leqslant 1. \tag{3.2}$$

Two events A and B are said to be *mutually exclusive* if the occurrence of one of them precludes the occurrence of the other. Rolling a three with a die and rolling a two are mutually exclusive events since both cannot occur simultaneously. If event A occurs $m(A)$ times and a mutually exclusive event B occurs $m(B)$ times in M trials, the probability that either A *or* B occurs is

$$\Pr(A \text{ or } B) = \lim_{M \to \infty} \frac{m(A) + m(B)}{M} = \Pr(A) + \Pr(B). \tag{3.3}$$

This equation is readily extended to any number of mutually exclusive events. For example, suppose the basic experiment is to count the number of gamma rays detected by a certain detector during a 1-sec interval. The probability that the number of detected photons $\mathbf{n}_d$ in this interval is precisely n is denoted by $\Pr(\mathbf{n}_d = n)$ or, more simply, $\Pr(n)$. The probability of detecting some different number n' is $\Pr(n')$. Then (3.3) says that

$$\Pr(n \text{ or } n') = \Pr(n) + \Pr(n'). \tag{3.4}$$

By extension, the probability that the number detected is less than or equal to some specified value N is given by

$$\Pr(\mathbf{n}_d \leqslant N) = \sum_{n=0}^{N} \Pr(n). \tag{3.5}$$

Since *some* number of photons (possibly zero) must be detected, the condition $N \to \infty$ represents the certain event, and we obtain an important normalization condition:

$$\Pr(\mathbf{n}_d < \infty) = \sum_{n=0}^{\infty} \Pr(n) = 1. \tag{3.6}$$

In the above example, the number of detected photons $\mathbf{n}_d$ is called a *random variable*. Basically, a random variable is nothing more than a numerical representation of the set of all possible outcomes of an experiment. One of those outcomes is called a *sample value* of the random variable. In the case of $\mathbf{n}_d$, the *domain* of the random variable is the set of nonnegative integers.

More generally, the domain is the set of values that can be assumed by the random variable. Other examples can easily be constructed where the domain consists of all integers, all real numbers, or even vectors or complex numbers. We shall use sans-serif type to denote a random variable in all cases. Sample values of the random variable will be in italic type for scalars and boldface roman type for vectors.

If we are interested in two or more random variables simultaneously, a *joint probability* is required. To make the discussion concrete, let us again consider a photon-counting experiment, this time with two detectors. The number counted by detector one during a certain time interval is denoted by the random variable $\mathbf{n}_1$, while the number counted by detector two during the *same* interval is the random variable $\mathbf{n}_2$. The joint probability $\Pr(\mathbf{n}_1 = n, \mathbf{n}_2 = m)$ or, more simply, $\Pr(n, m)$, is defined as the relative frequency of occurrence of the composite event in which $\mathbf{n}_1$ is precisely n *and* $\mathbf{n}_2$ is precisely m. Equation (3.2) is still applicable for this composite event and may be restated as

$$0 \leqslant \Pr(n, m) \leqslant 1. \tag{3.7}$$

Furthermore, the certain event consists of allowing all possible combinations of n and m, each combination being mutually exclusive with all other combinations. Therefore, the analog of (3.6) is

$$\sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \Pr(n, m) = 1. \tag{3.8}$$

In terms of the joint probability, the simple probability that detector one detects n photons, irrespective of the number detected by detector two, is

$$\Pr(\mathbf{n}_1 = n) = \sum_{m=0}^{\infty} \Pr(n, m). \tag{3.9}$$

[Note that the simplified notation $\Pr(n)$ could be misleading here since it could also be interpreted as $\Pr(\mathbf{n}_2 = n)$.]

Another useful kind of probability is the *conditional probability*. In the two-detector photon-counting experiment, we may define a quantity $\Pr(\mathbf{n}_1 = n | \mathbf{n}_2 = m)$ [or $\Pr(n|m)$] as the conditional probability that the random variable $\mathbf{n}_1$ takes on the value n when it is known that $\mathbf{n}_2$ takes on the value m. Conditional probability may still be interpreted as a relative frequency, but the number of trials M in the denominator of (3.1) must now include only those trials in which $\mathbf{n}_2$ happened to equal m. The numerator is the number of times that $\mathbf{n}_1$ was equal to n and $\mathbf{n}_2$ was equal to m, just as it would be in a joint probability. The difference between a joint and a conditional probability is thus in the denominator of the relative-frequency definition, not in the numerator. It follows that the joint and conditional

probabilities are related by

$$\Pr(\mathbf{n}_1 = n, \mathbf{n}_2 = m) = \Pr(\mathbf{n}_1 = n | \mathbf{n}_2 = m) \Pr(\mathbf{n}_2 = m), \tag{3.10}$$

or, in simplified form,

$$\Pr(n, m) = \Pr(n|m) \Pr(m). \tag{3.11}$$

Two random variables like $\mathbf{n}_1$ and $\mathbf{n}_2$ are said to be *statistically independent* if the value taken on by one of them has no influence on the value assumed by the other. In that case the conditional probability must reduce to a simple probability,

$$\Pr(\mathbf{n}_1 = n | \mathbf{n}_2 = m) = \Pr(\mathbf{n}_1 = n). \tag{3.12}$$

By (3.10), this requires that

$$\Pr(\mathbf{n}_1 = n, \mathbf{n}_2 = m) = \Pr(\mathbf{n}_1 = n) \Pr(\mathbf{n}_2 = m), \tag{3.13}$$

or

$$\Pr(n, m) = \Pr(n) \Pr(m). \tag{3.14}$$

Statistical independence between two random variables thus means that their joint probability is factorable into two simple probabilities.

To this point, we have considered only discrete random variables, i.e., random variables whose domain can be put into one-to-one correspondence with some set of integers (in the case of $\mathbf{n}_1$ and $\mathbf{n}_2$, the set of all nonnegative integers). It is often useful, however, to consider random variables defined on a continuous domain, or continuous random variables for short. Consider, for example, the random variable $\mathbf{x}$ representing points along a line extending from $-\infty$ to $+\infty$. It makes no sense to speak of the probability that $\mathbf{x}$ takes on some specific value, say x_1, because there are a nondenumerably infinite number of such values. The probability of observing precisely the specified one is thus vanishingly small. Nevertheless the general concept of probability is still useful. We can, for example, consider the probability that $\mathbf{x}$ assumes a value in the range $x_1 - \frac{1}{2}\Delta x_1 < \mathbf{x} \leqslant x_1 + \frac{1}{2}\Delta x_1$. If Δx_1 is small enough, we would expect this probability to be proportional to Δx_1, and it can also depend on x_1 itself. We can thus write, rather generally,

$$\Pr(x_1 - \tfrac{1}{2}\Delta x_1 < \mathbf{x} \leqslant x_1 + \tfrac{1}{2}\Delta x_1) = \mathrm{pr}(x_1)\Delta x_1. \tag{3.15}$$

The proportionality factor $\mathrm{pr}(x_1)$ is called the *probability density function* for the random variable $\mathbf{x}$. There is an important difference between a *probability*, denoted $\Pr(\ )$, and a probability density function, denoted $\mathrm{pr}(\ )$ The latter may indeed be regarded as a mathematical function of its argument. When necessary, as when more than one random variable is involved in a given problem, we shall distinguish the corresponding probability density functions by appropriate subscripts.

With the aid of $\mathrm{pr}(x_1)$, more general probabilities can readily be calculated. For example, the probability that $\mathbf{x}$ falls in the finite range $a < \mathbf{x} < b$ may be found by dividing the range into small, mutually exclusive elements of width Δx_1. By generalization of (3.4), we then have

$$\Pr(a < \mathbf{x} \leqslant b) = \int_a^b \mathrm{pr}(x_1)\,dx_1. \tag{3.16}$$

Since $\mathbf{x}$ is constrained to lie in the interval $(-\infty, \infty)$, we must also have, by analogy to (3.6),

$$\Pr(-\infty < \mathbf{x} < \infty) = \int_{-\infty}^{\infty} \mathrm{pr}(x_1)\,dx_1 = 1. \tag{3.17}$$

The analog of (3.5) is also useful. By setting $a = -\infty$ and $b = X$ in (3.16) we find

$$\Pr(\mathbf{x} \leqslant X) = \int_{-\infty}^{X} \mathrm{pr}(x_1)\,dx_1. \tag{3.18}$$

This quantity, $\Pr(\mathbf{x} \leqslant X)$, regarded as a function of X, is called the *cumulative probability distribution function** for the random variable $\mathbf{x}$. If $\Pr(\mathbf{x} \leqslant X)$ is known, $\mathrm{pr}(x_1)$ is readily found from it since

$$\frac{\partial}{\partial X}\Pr(\mathbf{x} \leqslant X) = \frac{\partial}{\partial X}\int_{-\infty}^{X} \mathrm{pr}(x_1)\,dx_1 = \mathrm{pr}(X). \tag{3.19}$$

Since $\mathrm{pr}(X)$ cannot be negative [see (3.15)], $\Pr(\mathbf{x} \leqslant X)$ must be a monotonically increasing function of X.

When more than one continuous random variable is involved, joint and conditional forms of both the probability density function and the cumulative probability distribution function may be defined by a straightforward extension of their discrete counterparts.

Even if the random variables of interest are discrete, the continuous formalism is still applicable if we allow the probability density functions to involve Dirac delta functions. For example, suppose that the random variable $\mathbf{x}$ can take on only the discrete values x_n $(n = 1, 2, \ldots, N)$ and that

$$\Pr(\mathbf{x} = x_n) = P_n. \tag{3.20}$$

Then we can define a probability density function $\mathrm{pr}(x')$ by

$$\mathrm{pr}(x') = \sum_{n=1}^{N} P_n \delta(x' - x_n). \tag{3.21}$$

* Many texts refer to this quantity as the *distribution function*, but we shall retain the word *cumulative* to avoid confusion with specific probabilities, such as for Poisson random variables, which are also called distributions.

To see that (3.21) is indeed consistent with (3.20), let us integrate it over some vanishingly small range:

$$\int_{x-\varepsilon}^{x+\varepsilon} \mathrm{pr}(x')\,dx' = \Pr(x - \varepsilon < \mathbf{x} \leqslant x + \varepsilon) = \begin{cases} P_n & \text{if } x = x_n \\ 0 & \text{otherwise,} \end{cases} \tag{3.22}$$

which must be correct since $\mathbf{x}$ can fall in the range $(x - \varepsilon, x + \varepsilon)$ only if one of its discrete values lies in that range.

The reader may wish to verify that if $\mathbf{x}$ is constrained to discrete values, $\Pr(\mathbf{x} \leqslant X)$ is discontinuous but finite and monotonically increasing everywhere.

3.1.2 Expectation Values, Moments, and the Characteristic Function

Consider a discrete random variable $\mathbf{x}$ that can have the values x_i $(i = 1, 2, \ldots, N)$. Suppose a large number of measurements of $\mathbf{x}$ (or equivalently, experiments of which $\mathbf{x}$ is the outcome) are performed, and that the value x_i occurs $m(x_i)$ times. Then the arithmetic average value of $\mathbf{x}$ is given by

$$\langle \mathbf{x} \rangle_M = \frac{1}{M}\sum_{i=1}^{N} m(x_i)x_i, \tag{3.23}$$

where M is the total number of measurements, i.e.,

$$M = \sum_{i=1}^{N} m(x_i). \tag{3.24}$$

But, from the definition of probability in (3.1), $m(x_i)/M$ approaches $\Pr(\mathbf{x} = x_i) = \Pr(x_i)$ as $M \to \infty$. Therefore, in this limit,

$$\lim_{M\to\infty} \langle \mathbf{x} \rangle_M \equiv \langle \mathbf{x} \rangle = \sum_{i=1}^{N} x_i \Pr(x_i). \tag{3.25}$$

The natural generalization of this equation to a continuous random variable of domain $(-\infty, \infty)$ is

$$\langle \mathbf{x} \rangle = \int_{-\infty}^{\infty} x' \,\mathrm{pr}(x')\,dx'. \tag{3.26}$$

The quantity $\langle \mathbf{x} \rangle$ is variously called *the mean, the ensemble average, the stochastic average, the expected value*, or *the expectation value* of $\mathbf{x}$. It is often denoted by $E\{\mathbf{x}\}$ or $\bar{x}$.

The same line of reasoning can be applied to determine the expectation value of any power of $\mathbf{x}$. The random variable $\mathbf{y} = \mathbf{x}^n$ takes on the value x_i^n whenever $\mathbf{x} = x_i$. Since this condition obtains $m(x_i)$ times in M measurements,

we can write

$$\langle \mathbf{x}^n \rangle = \sum_{i=1}^{N} x_i^n \Pr(x_i) \tag{3.27}$$

or, in the continuous case,

$$\langle \mathbf{x}^n \rangle = \int_{-\infty}^{\infty} x'^n \operatorname{pr}(x')\,dx'. \tag{3.28}$$

Any function of $\mathbf{x}$, at least any one sufficiently well behaved to admit of a power-series expansion, can be treated similarly, yielding

$$\langle f(\mathbf{x}) \rangle = \sum_{i=1}^{N} f(x_i) \Pr(x_i) \tag{3.29}$$

or

$$\langle f(\mathbf{x}) \rangle = \int_{-\infty}^{\infty} f(x') \operatorname{pr}(x')\,dx'. \tag{3.30}$$

The quantity $\langle \mathbf{x}^n \rangle$ is called the *nth moment* of $\mathbf{x}$. Also of interest is the *nth central moment* defined by

$$\langle (\mathbf{x} - \bar{x})^n \rangle = \sum_{i=1}^{N} (x_i - \bar{x})^n \Pr(x_i) \tag{3.31}$$

or

$$\langle (\mathbf{x} - \bar{x})^n \rangle = \int_{-\infty}^{\infty} (x' - \bar{x})^n \operatorname{pr}(x')\,dx', \tag{3.32}$$

where

$$\bar{x} \equiv \langle \mathbf{x} \rangle. \tag{3.33}$$

In particular, the *second central moment* or *variance* is given by

$$\sigma_x^2 = \langle (\mathbf{x} - \bar{x})^2 \rangle = \langle \mathbf{x}^2 \rangle - \langle \mathbf{x} \rangle^2. \tag{3.34}$$

The second form follows because $\langle \mathbf{x}\,\bar{x} \rangle = \bar{x}\langle \mathbf{x} \rangle = \bar{x}^2 = \langle \mathbf{x} \rangle^2$. The positive square root of the variance is called the *standard deviation of* $\mathbf{x}$, denoted by σ_x, and is the root-mean-square (rms) value of the fluctuations of $\mathbf{x}$ about its mean.

When more than one random variable is involved in a problem, expectation values and moments must be found from the joint probability functions. For example, if $\mathbf{x}$ and $\mathbf{y}$ are continuous random variables of domain $(-\infty, \infty)$, a joint moment or cross moment of order $n + m$ is given by

$$\langle \mathbf{x}^n \mathbf{y}^m \rangle = \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy'\, x'^n y'^m \operatorname{pr}(x', y'). \tag{3.35}$$

The joint central moment $\langle (\mathbf{x} - \bar{x})(\mathbf{y} - \bar{y}) \rangle$ is called the *covariance* of $\mathbf{x}$ and $\mathbf{y}$.

If $\mathbf{x}$ and $\mathbf{y}$ are statistically independent, $\operatorname{pr}(x', y') = \operatorname{pr}(x')\operatorname{pr}(y')$ and hence

$$\langle \mathbf{x}^n \mathbf{y}^m \rangle = \langle \mathbf{x}^n \rangle \langle \mathbf{y}^m \rangle. \tag{3.36}$$

On the other hand, if $\langle \mathbf{xy} \rangle = \langle \mathbf{x} \rangle \langle \mathbf{y} \rangle$, $\mathbf{x}$ and $\mathbf{y}$ are said to be *uncorrelated*. This is a weaker condition than statistical independence since it could happen that $\langle \mathbf{xy} \rangle = \langle \mathbf{x} \rangle \langle \mathbf{y} \rangle$, but $\langle \mathbf{x}^n \mathbf{y}^m \rangle \neq \langle \mathbf{x}^n \rangle \langle \mathbf{y}^m \rangle$ if $n \neq 1$ or $m \neq 1$.

A particularly interesting statistical average of a function of the random variable $\mathbf{x}$ is the quantity $M_x(\lambda)$ defined by

$$M_x(\lambda) = \langle \exp(-2\pi i \lambda \mathbf{x}) \rangle. \tag{3.37}$$

If $\mathbf{x}$ is a continuous random variable of domain $(-\infty, \infty)$, then from (3.30)

$$M_x(\lambda) = \int_{-\infty}^{\infty} dx' \operatorname{pr}(x') \exp(-2\pi i \lambda x'). \tag{3.38}$$

In other words, $M_x(\lambda)$, which is usually called the *characteristic function*, is the Fourier transform of the probability density function $\operatorname{pr}(x')$.

Another name for $M_x(\lambda)$ is the *moment-generating function* [although many books reserve this term for $\langle \exp(\lambda \mathbf{x}) \rangle$]. The reason for this designation is that the nth moment of $\mathbf{x}$ is related to the nth derivative of $M_x(\lambda)$ as follows:

$$\begin{aligned} \left(\frac{d^n M_x(\lambda)}{d\lambda^n} \right)_{\lambda = 0} &= (-2\pi i)^n \int_{-\infty}^{\infty} \operatorname{pr}(x') x'^n\,dx' \\ &= (-2\pi i)^n \langle \mathbf{x}^n \rangle. \end{aligned} \tag{3.39}$$

This equation also shows that the characteristic function, and hence the probability density function, is completely determined by the moments, because the left-hand side is just the coefficient of the nth term in a Taylor series expansion of $M_x(\lambda)$.

As an example of the use of the characteristic function, let us consider the random variable $\mathbf{z} = \mathbf{x} + \mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y}$ are statistically independent random variables. The characteristic function for $\mathbf{z}$ is

$$M_z(\lambda) = \langle \exp[-2\pi i \lambda (\mathbf{x} + \mathbf{y})] \rangle. \tag{3.40}$$

Because of the statistical independence, this becomes

$$\begin{aligned} M_z(\lambda) &= \langle \exp(-2\pi i \lambda \mathbf{x}) \rangle \langle \exp(-2\pi i \lambda \mathbf{y}) \rangle \\ &= M_x(\lambda) M_y(\lambda). \end{aligned} \tag{3.41}$$

Now we can use the fact that $M_z(\lambda)$ is the Fourier transform of $\operatorname{pr}_z(z')$, together with the convolution theorem (B.52) to obtain the important result that the probability density function for $\mathbf{z}$ is the convolution of the corresponding functions for $\mathbf{x}$ and $\mathbf{y}$. Adding subscripts for clarity, we may write

this result

$$\mathrm{pr}_z(z') = \mathrm{pr}_x(z') * \mathrm{pr}_y(z') = \int_{-\infty}^{\infty} \mathrm{pr}_x(x')\,\mathrm{pr}_y(y' = z' - x')\,dx'. \tag{3.42}$$

Further use will be made of this result in Section 3.2.

3.1.3 Random Processes

The behavior of a random variable is often influenced by one or more independent parameters such as time or position. When the dependence on these parameters is important, the random variable is called a *random process*. For example, the instantaneous count rate from a detector viewing a radioisotopic gamma-ray source is a random variable that fluctuates with time (see Fig. 3.1). Furthermore, even the *average* count rate may vary with time if the observation time is long compared to the half-life of the source or if the source-detector distance is changing with time. In such a case the relative-frequency definition of probability encounters difficulties since it is based on repeated or time-sequential measurements. An implicit assumption of this definition is that there are no time-dependent effects, that every repetition satisfies the same conditions.

This difficulty can be circumvented by means of a gedankenexperiment in which we imagine that we have an ensemble of a large number of identical apparatuses with all parameters the same. In our example, the ensemble consists of a large number of radioactive sources, all of the same composition and activity, and all viewed by identical detectors in identical configurations. Then, at some instant t_1, we activate all the counters and count for one second, thereby simultaneously obtaining a large number of values for the random variable $\mathbf{n}_d(t_1)$ (See Fig. 3.1). Of course, in spite of our best efforts, these values for $\mathbf{n}_d(t_1)$ will not all be the same. The process is still inherently statistical because we cannot control, even in a gedankenexperiment, when a given nucleus decays and in what direction it emits its gamma ray. However, the important point is that we now have, in principle, a set of measurements with which to define relative frequency, probability, moments, and the whole panoply. Averages formed in this way are called *ensemble averages*; they are characteristic of $\mathbf{n}_d$ at one particular time t_1 and say nothing at all about how the system evolves in time.

To trace the time evolution of the ensemble, we must repeat the entire set of measurements at a sequence of times $t_1, t_2, \ldots$. Each set of measurements $\mathbf{n}_d(t_i)$ can then be used to define a separate probability function $\Pr[\mathbf{n}_d(t_i) = n]$. In other words, there is one probability function for $\mathbf{n}_d(t_1)$,

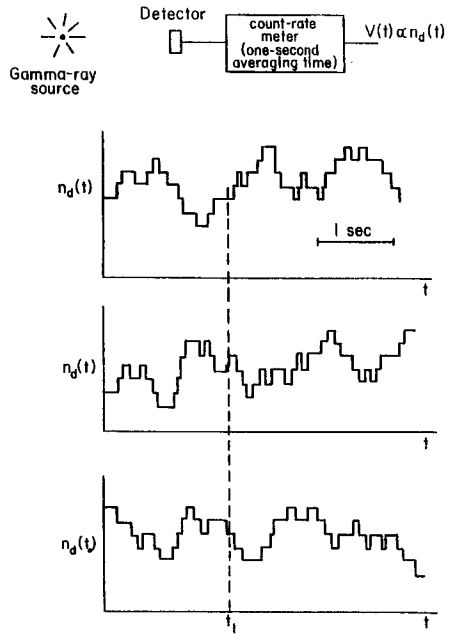

Fig. 3.1 Illustration of a simple gamma-ray counting experiment using a count-rate meter. The output of this device is a voltage $\mathbf{V}(t)$ proportional to the number of photons detected during the previous second, which is the random variable $\mathbf{n}_d(t)$. Also shown are three sample functions of this random variable, representing three different members of an ensemble of identical apparatuses. Averages "along the process," i.e., along the time axis of any one sample function, are called time averages. Averages "across the process," averaging different sample functions at one instant t_1, are called ensemble averages.

another for $\mathbf{n}_d(t_2)$, etc.; each $\mathbf{n}_d(t_i)$ is treated as a separate, though not necessarily independent, random variable.

We are now faced with a different kind of problem: How are the different random variables $\mathbf{n}_d(t_i)$ related to each other? A complete answer to this question requires the determination of the joint probability function relating all of the random variables, $\Pr[\mathbf{n}_d(t_i) = n_1,\ \mathbf{n}_d(t_2) = n_2, \ldots]$. However, since the t_i can be arbitrarily close, there can be a nondenumerably infinite number of $\mathbf{n}_d(t_i)$, making evaluation of the joint probability hopeless. Fortunately, we can adequately characterize the system for most purposes if we can determine a simpler joint probability, the one obtained by considering just two of the $\mathbf{n}_d(t_i)$ at a time. If we know $\Pr[\mathbf{n}_d(t_i) = n, \mathbf{n}_d(t_j) = m]$, we have enough information to calculate moments like $\langle[\mathbf{n}_d(t_i)]^k[\mathbf{n}_d(t_j)]^l\rangle$. Except in rare circumstances, we do not need to know more. In fact, in most cases

we can get by with still less and merely calculate a few of the lower-order joint moments without even finding this pairwise joint probability.

One important joint moment is the one for which $k = l = 1$, i.e., $\langle \mathbf{n}_d(t_i)\mathbf{n}_d(t_j)\rangle$. This quantity, regarded as a function of t_i and t_j, is called the *autocorrelation function* for the random process $\mathbf{n}_d(t)$. In terms of the pairwise joint probability, it is given by

$$R_{n_d}(t_i, t_j) \equiv \langle \mathbf{n}_d(t_i)\mathbf{n}_d(t_j)\rangle = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} nm \Pr[\mathbf{n}_d(t_i) = n, \mathbf{n}_d(t_j) = m]. \tag{3.43}$$

More generally, we may consider a continuous random process $\mathbf{x}(t)$ [where for each t_i, $\mathbf{x}(t_i)$ ranges over the continuous domain from $-\infty$ to ∞]. Then the autocorrelation function is given by

$$R_x(t_i, t_j) = \langle \mathbf{x}(t_i)\mathbf{x}(t_j)\rangle = \int_{-\infty}^{\infty} dx'(t_i) \int_{-\infty}^{\infty} dx'(t_j) x'(t_i)x'(t_j) \operatorname{pr}[x'(t_i), x'(t_j)]. \tag{3.44}$$

If $\mathbf{x}(t_i)$ is a complex random process, it is conventional to define the autocorrelation by $\langle \mathbf{x}(t_i)\mathbf{x}^*(t_j)\rangle$, with the asterisk denoting complex conjugation. In this case the integral in (3.44) becomes four-dimensional, spanning the real and imaginary parts of both $x'(t_i)$ and $x'(t_j)$.

So far, we have discussed autocorrelations in which the independent parameter was the time t, but other parameters can also be used. For example, in Chapter 10 spatial autocorrelations of the form $\langle \mathbf{x}(\mathbf{r}_i)\mathbf{x}(\mathbf{r}_j)\rangle$ will receive considerable attention. However, for the nonce, we shall continue to use t as the parameter in order to avoid the complexity of vector notation.

In many problems, the autocorrelation function for some random process $\mathbf{x}(t)$ turns out not to depend on the two time arguments t_i and t_j separately but only on their difference $t_i - t_j$. Such processes are said to be *stationary*, and the autocorrelation function may be written

$$R_x(\tau) = \langle \mathbf{x}(t_i)\mathbf{x}(t_i - \tau)\rangle, \tag{3.45}$$

where $\tau = t_i - t_j$. (Some authors prefer to call this condition wide-sense stationarity; narrow-sense or strict stationarity is then taken as the stronger condition that the pairwise joint probability density function be a function of only $t_i - t_j$.) Since $\mathbf{x}(t)$ is stationary, $R_x(\tau)$ must be unchanged if we replace t_i with $t_i + \tau$ in (3.45). Therefore, for a real stationary random process, we must have $R_x(\tau) = R_x(-\tau)$. [More generally, if $\mathbf{x}(t)$ is stationary but complex, $R_x(\tau) = R_x^*(-\tau)$.] It can also be shown that $R_x(\tau)$ is a maximum at the origin, i.e., $|R_x(\tau)| \leq R_x(0)$. From the definition (3.45), $R_x(0)$ is seen to be the second moment of $\mathbf{x}(t)$.

In the gedankenexperiment with the ensemble of sources and detectors, the output from any one detector as a function of time is called a *sample function* of the random process $\mathbf{n}_d(t)$. Now, in real life, we are unlikely to have anything approaching an ensemble of detectors at our disposal; more likely, we shall have only one. The question then is, What can be learned about the random process from a measurement of one of its sample functions? Certainly averages can still be defined, only now they must be time averages rather than statistical or ensemble averages. A finite-duration time average of the sample function can be defined as

$$\langle n_d(t)\rangle_T = \frac{1}{T}\int_{t-(T/2)}^{t+(T/2)} n_d(t')\,dt'. \tag{3.46}$$

The time average of a random process may not exist or it may be different for different sample functions. Even if these problems do not crop up, the time average may not be a useful indicator of the statistical average, since the latter may itself be a function of time. Nevertheless, there are some conditions under which the time average should equal the statistical average. Returning to our photon-counting example, if the source has a very long half-life and the geometry of the apparatus is constant, there is no reason to believe that a 1-sec count performed now and one performed an hour from now will be any different statistically; the probability functions for the two counts should be the same. Furthermore, there is no reason to expect the outcome of the first count to have any influence on the second; the two counts are statistically independent and the joint probability function factorizes. Under these circumstances the two counts taken one hour apart could equally well have been taken simultaneously on two independent apparatuses, i.e., they can be regarded as two members of an ensemble. In this case the ensemble average and the time average* should give the same answer for the mean counting rate. In general, when statistical averages can be replaced by time averages, the process is said to be *ergodic*. An ergodic process must be stationary, but stationarity does not guarantee ergodicity. Proving ergodicity is a difficult statistical problem that will not be pursued here.

If a process is ergodic, any reasonable function of the process is also ergodic. Therefore, moments can be calculated by time averaging as well as by ensemble averaging. By the same token, we may define a *temporal autocorrelation function* for a sample function of a real random process by

$$\mathscr{R}_x(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{t-(T/2)}^{t+(T/2)} x(t' + \tau)x(t')\,dt'. \tag{3.47}$$

If the process is ergodic, $\mathscr{R}_x(\tau)$ will be independent of t and equal to $R_x(\tau)$ defined in (3.45).

* Taken over a time long enough to give good statistical accuracy but short compared to the half-life.

The reader may show some consternation at this point since $\mathscr{R}_x(\tau)$ is the third distinctly different quantity we have called an autocorrelation function [cf. (3.45), (3.47), and (B.53)]. The only defense is that all three usages are thoroughly entrenched in the literature. *Caveat lector.*

3.1.4 Spectral Analysis

Throughout this book we shall make good use of frequency-domain representations of various physical quantities. It is natural to expect that the frequency domain will also prove useful with random variables. A straightforward extension of Fourier concepts into the stochastic realm is, however, soon met with formidable mathematical difficulties.

For example, a Fourier transform of a sample function $x(t)$ of a stationary random process might be defined by

$$X(\nu) = \int_{-\infty}^{\infty} x(t)\exp(-2\pi i\nu t)\,dt. \tag{3.48}$$

The problem with this approach is that the total energy in this signal, defined by

$$\int_{-\infty}^{\infty} |x(t)|^2\,dt = \int_{-\infty}^{\infty} |X(\nu)|^2\,d\nu, \tag{3.49}$$

must be infinite if the process is stationary. The random variable fluctuates about its mean with constant mean-square deviation and therefore does not possess a Fourier transform in the usual sense.

One way around this difficulty is to assume that the random process is periodic, and then to let the period approach infinity. A stationary random process is said to be periodic with period T if its autocorrelation function obeys

$$R_x(t+T) = R_x(t) \tag{3.50}$$

or, equivalently, if every sample function $x(t)$ is periodic with period T. Then we can expand $\mathbf{x}(t)$ in a Fourier series as

$$\mathbf{x}(t) = \sum_{n=-\infty}^{\infty} \mathbf{X}_n \exp(2\pi i n\nu_0 t), \tag{3.51}$$

where

$$\mathbf{X}_n = \frac{1}{T}\int_{-T/2}^{T/2} \mathbf{x}(t)\exp(-2\pi i n\nu_0 t)\,dt \tag{3.52}$$

and $\nu_0 = 1/T$. [Equations like (3.51) and (3.52), where there are random processes on both sides, are valid equations if they hold for all sample functions of the random process.]

The autocorrelation function can be similarly expanded:

$$R_x(\tau) = \sum_{k=-\infty}^{\infty} b_k \exp(2\pi i k\nu_0\tau), \tag{3.53a}$$

where

$$b_k = \frac{1}{T}\int_{-T/2}^{T/2} R_x(\tau)\exp(-2\pi i k\nu_0\tau)\,d\tau. \tag{3.53b}$$

It is now straightforward to compute the correlation properties of the Fourier coefficients $\mathbf{X}_n$. For $\mathbf{x}(t)$ real, we find

$$\begin{aligned}\langle \mathbf{X}_n\mathbf{X}_m^*\rangle &= \frac{1}{T^2}\int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt' \langle \mathbf{x}(t)\mathbf{x}(t')\rangle \exp[-2\pi i\nu_0(nt - mt')] \\ &= \frac{1}{T^2}\int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt'\, R_x(t'-t)\exp[-2\pi i\nu_0(nt - mt')] \\ &= \frac{1}{T^2}\int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt' \sum_{k=-\infty}^{\infty} b_k \exp[2\pi i\nu_0(k-n)t] \\ &\quad \times \exp[-2\pi i\nu_0(k-m)t'].\end{aligned} \tag{3.54}$$

However, the integral over t vanishes unless $k = n$, in which case it has the value T. Similarly, the integral over t' vanishes unless $k = m$, in which case it also has the value T. Hence

$$\langle \mathbf{X}_n\mathbf{X}_m^*\rangle = \begin{cases} b_n & \text{if } n = m \\ 0 & \text{if } n \neq m. \end{cases} \tag{3.55}$$

The Fourier coefficients $\mathbf{X}_n$ are therefore uncorrelated and simply related to the Fourier coefficients of $R_x(\tau)$.

Since b_n is just the average power* $\langle |\mathbf{X}_n|^2\rangle$ associated with the frequency $n\nu_0$, we are led to define a *power spectral density* function or *Wiener spectrum* by

$$S_x(\nu) = \sum_{n=-\infty}^{\infty} b_n\,\delta(\nu - n\nu_0). \tag{3.56}$$

By taking the Fourier transform of (3.53a), we see that $S_x(\nu)$ is simply the Fourier transform of $R_x(\tau)$:

$$S_x(\nu) = \int_{-\infty}^{\infty} d\tau\, R_x(\tau)\exp(-2\pi i\nu\tau). \tag{3.57}$$

* The term "power" is a carryover from electrical engineering. If $\mathbf{x}(t)$ is the voltage across a $1-\Omega$ resistor, $\langle |\mathbf{x}(t)|^2\rangle$ is the average power dissipated.

The least tortuous way of extending this discussion to nonperiodic random processes is just to regard (3.57), which is often called the *Wiener–Khinchin theorem*, as the definition of the power spectral density. We shall see in Section 3.4 that the general definition, (3.57), is consistent with the strictly periodic definition, (3.56), in the sense that the integral of $S_x(\nu)$ over some small frequency range $\Delta\nu$ still represents the average power in that range. In support of this interpretation, it may be noted that $S_x(\nu)$ cannot be negative.

3.2 GAUSSIAN RANDOM VARIABLES

The discussion of random variables to this point has been rather abstract; no specific functional forms have been used for either the cumulative probability distribution function or the probability density functions. Although a great variety of density functions appear in the literature, virtually all statistical problems in radiographic imaging involve just two types of random variables—the Poisson random variable treated in the next section and the Gaussian random variable discussed here.

3.2.1 The Normal Distribution

A continuous random variable $\mathbf{x}$ of domain $(-\infty, \infty)$ is said to be *Gaussian* or to have a *normal distribution* if its probability density function is a Gaussian of the form

$$\text{pr}(x') = (2\pi\sigma_x^2)^{-1/2} \exp[-(x' - \bar{x})^2/2\sigma_x^2], \tag{3.58}$$

which is plotted in Fig. 3.2. Direct integration will confirm that this function is properly normalized and obeys (3.17). It is also easy to show that $\bar{x}$ is, as the notation implies, the mean of $\mathbf{x}$, and σ_x^2 is its variance, i.e.,

$$\bar{x} = \int_{-\infty}^{\infty} x' \,\text{pr}(x')\, dx', \tag{3.59}$$

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x' - \bar{x})^2 \,\text{pr}(x') dx'. \tag{3.60}$$

It is an important property of Gaussian random variables that all higher moments are determined by $\bar{x}$ and σ_x, since these are the only parameters that appear in pr(x'). A repeated integration by parts shows, for example, that the kth central moment is

$$\langle (\mathbf{x} - \bar{x})^k \rangle = \begin{cases} 1 \cdot 3 \cdot \ldots \cdot (k-1)\sigma_x^k, & k \text{ even} \\ 0, & k \text{ odd.} \end{cases} \tag{3.61}$$

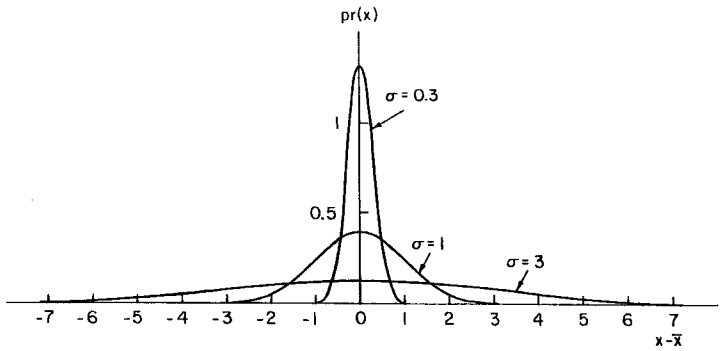

Fig. 3.2 Plots of three different Gaussian probability density functions with different variances.

Since pr(x') is a Gaussian, its Fourier transform, the characteristic function, is also a Gaussian. By Eqs. (B.12), (B.14), and (B.37),

$$M_x(\lambda) = \mathcal{F}_1\{\text{pr}(x')\} = \exp[-2\pi i\lambda\bar{x} - 2\pi^2\sigma_x^2\lambda^2]. \tag{3.62}$$

The reader can verify this form for $M_x(\lambda)$ by differentiating it and using (3.39) to show that the moments given in (3.61) are obtained.

The cumulative probability distribution function $\text{Pr}(\mathbf{x} < X)$ comes up often in statistical problems. For the special case of a Gaussian random variable of zero mean and unit variance ($\bar{x} = 0$, $\sigma_x^2 = 1$), $\text{Pr}(\mathbf{x} < X)$ is related to the *error function* erf(u) by

$$\text{Pr}(\mathbf{x} < X) = (2\pi)^{-1/2} \int_{-\infty}^{X} \exp\left(\frac{-x^2}{2}\right) dx = \frac{1}{2}\left[1 + \text{erf}\left(\frac{X}{\sqrt{2}}\right)\right], \tag{3.63a}$$

$$\text{erf}(u) \equiv \frac{2}{\sqrt{\pi}} \int_0^u \exp(-t^2)\, dt. \tag{3.63b}$$

Tables of the error function may be found in most books on statistics. Situations where $\bar{x} \neq 0$ or $\sigma_x^2 \neq 1$ may be handled by a change of variables.

Two random variables $\mathbf{x}$ and $\mathbf{y}$ are said to be *jointly normal* or *bivariate normal* if their joint probability density function is of the form

$$\text{pr}(x', y') = [4\pi^2\sigma_x^2\sigma_y^2(1-\gamma^2)]^{-1/2} \times \exp\left[-\frac{1}{2(1-\gamma^2)}\left(\frac{(x'-\bar{x})^2}{\sigma_x^2} - \frac{2\gamma(x'-\bar{x})(y'-\bar{y})}{\sigma_x\sigma_y} + \frac{(y'-\bar{y})^2}{\sigma_y^2}\right)\right]. \tag{3.64}$$

The new parameter γ is called the *correlation coefficient* and is a normalized measure of the correlation between $\mathbf{x}$ and $\mathbf{y}$. If $\gamma = 0$, $\mathrm{pr}(x', y')$ factors into two normal distributions and $\mathbf{x}$ and $\mathbf{y}$ are statistically independent. When $\gamma \neq 0$, its significance can be appreciated by calculating the covariance of $\mathbf{x}$ and $\mathbf{y}$, given (after a tedious integration) by

$$\langle (\mathbf{x} - \bar{x})(\mathbf{y} - \bar{y}) \rangle = \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' (x' - \bar{x})(y' - \bar{y}) \,\mathrm{pr}(x', y')$$
$$= \gamma \sigma_x \sigma_y = \gamma [\sigma_x^2 \sigma_y^2]^{1/2}. \tag{3.65}$$

The correlation coefficient is thus the covariance normalized by the geometric mean of the two variances.

For most of the problems in this book, we shall be concerned with such random processes as $\mathbf{x}(t)$. Sometimes, but certainly not always, it will turn out that $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$ may be considered to be jointly normal random variables with some degree of correlation between them. The correlation coefficient is related to the autocorrelation function for the process $R_x(t_1, t_2)$ [see (3.43)] by

$$\gamma(t_1, t_2) = \frac{\langle [\mathbf{x}(t_1) - \bar{x}(t_1)][\mathbf{x}(t_2) - \bar{x}(t_2)] \rangle}{\sigma_{x(t_1)} \sigma_{x(t_2)}}$$
$$= [\sigma_{x(t_1)} \sigma_{x(t_2)}]^{-1} [R_x(t_1, t_2) - \bar{x}(t_1)\bar{x}(t_2)]. \tag{3.66}$$

For the important special case of stationary statistics,

$$\bar{x}(t_1) = \bar{x}(t_2) = \bar{x}, \qquad \sigma_{x(t_1)} = \sigma_{x(t_2)} = \sigma_x, \qquad R_x(t_1, t_2) = R_x(t_1 - t_2),$$

and

$$\gamma(t_1, t_2) = \sigma_x^{-2} [R_x(t_1 - t_2) - \bar{x}^2]. \tag{3.67}$$

It is worth noting that no new restrictions have been placed on $R_x(\tau)$ in this section. A Gaussian random process may have a great variety of temporal behaviors. Some statement about the autocorrelation function, or equivalently, the power spectral density, must be included to fully characterize the process.

3.2.2 The Central-Limit Theorem

Although Gaussian probability densities are comparatively easy to manipulate, our interest in them is based on more than just mathematical convenience. More important, they occur with great regularity in real physical problems because of an important principle known as the *central-limit theorem.*

In many physical problems, the observed random variable is really the sum of a large number N of other independent random variables. In this case, the central-limit theorem states that the probability density of the observed variable approaches a Gaussian as N tends to infinity, *regardless of the densities of the constituent random variables.*

This remarkable theorem is most easily understood by use of the characteristic function introduced in Section 3.1.2. The treatment here is an extension of the arguments that led to (3.42). Let $\mathbf{z}$ denote the observed random variable and $\mathbf{x}_j$ be the jth constituent variable, so that

$$\mathbf{z} = \sum_{j=1}^{N} \mathbf{x}_j. \tag{3.68}$$

The characteristic function for $\mathbf{z}$ is

$$M_z(\lambda) = \langle \exp(-2\pi i \lambda \mathbf{z}) \rangle = \left\langle \exp\left(-2\pi i \lambda \sum_{j=1}^{N} \mathbf{x}_j \right) \right\rangle. \tag{3.69}$$

Since the $\mathbf{x}_j$ are assumed to be independent, $M_z(\lambda)$ becomes

$$M_z(\lambda) = \prod_{j=1}^{N} \langle \exp(-2\pi i \lambda \mathbf{x}_j) \rangle = \prod_{j=1}^{N} M_{x_j}(\lambda). \tag{3.70}$$

Since each characteristic function is the Fourier transform of the corresponding probability density function, we have

$$\mathrm{pr}_z(z') = \mathrm{pr}_1(z') * \mathrm{pr}_2(z') * \cdots * \mathrm{pr}_N(z'). \tag{3.71}$$

where $\mathrm{pr}_j(z')$ is the probability density for $\mathbf{x}_j$ evaluated at the point z'.

The important qualitative point about this result is that each of the functions $\mathrm{pr}_j(z')$ is positive definite, as all probability densities must be. As discussed in Section 2.4, convolution with a positive-definite function is a smoothing operation; therefore $\mathrm{pr}_z(z')$ becomes progressively smoother as N increases, regardless of the functional form for $\mathrm{pr}_j(z')$.

Of course, it is not yet obvious that the final smooth form for $\mathrm{pr}_z(z')$ is, in fact, Gaussian, although certainly a Gaussian is a very smooth function. The graphical example in Fig. 3.3 will lend some credence to the contention that $\mathrm{pr}_z(z')$ approaches a Gaussian, but a more formal derivation is also desirable. A rigorous derivation, valid for all sorts of pathological density functions, is actually quite difficult. However, we shall sketch the proof for the special case where all of the pr_j have the same functional form and are not terribly ill behaved. Our treatment parallels that of Davenport and Root (1958), to which the reader is referred for a more rigorous justification of the approximations used.

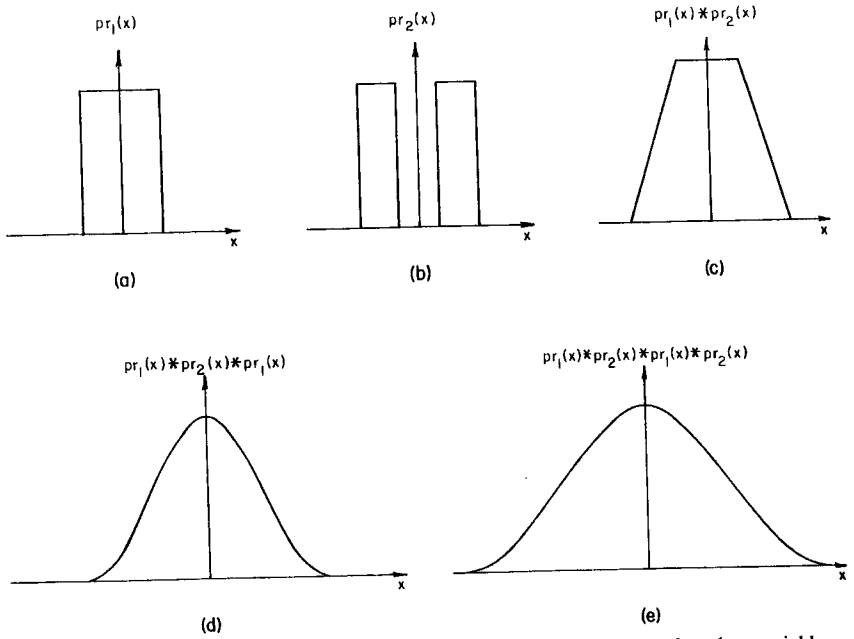

Fig. 3.3 Illustration of the central limit theorem. We consider a sum of random variables $\mathbf{x}_j$, where $\mathrm{pr}_j(x)$ has the form shown in (a) for j odd and in (*b*) for j even. Density functions for $\mathbf{x}_1 + \mathbf{x}_2$, $\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3$, and $\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 + \mathbf{x}_4$ are shown in (c), (d), and (e), respectively. Note that even though the constituent densities were decidedly non-Gaussian, (e) is already an excellent approximation to a Gaussian.

It is most convenient to work with normalized random variables having zero mean and unit variance. Therefore, we define

$$\boldsymbol{\chi}_j = (\mathbf{x}_j - \bar{x})/\sigma_x, \tag{3.72}$$

$$\boldsymbol{\zeta} = (\mathbf{z} - \bar{z})/\sigma_z. \tag{3.73}$$

Note $\bar{x}$ and σ_x^2 do not require the subscript j, since all of the $\mathbf{x}_j$ have the same statistics. It is easy to verify that

$$\langle \boldsymbol{\chi}_j \rangle = 0, \qquad \sigma_\chi^2 = 1, \tag{3.74}$$

$$\langle \boldsymbol{\zeta} \rangle = 0, \qquad \sigma_\zeta^2 = 1. \tag{3.75}$$

Now the mean value of a sum of statistically independent random variables is just the sum of the mean values of the individual variables, and similarly, the variance of the sum is the sum of the variances. Therefore,

$$\bar{z} = N\bar{x}, \qquad \sigma_z^2 = N\sigma_x^2. \tag{3.76}$$

These relations enable us to restate (3.68) in terms of the normalized variables $\boldsymbol{\zeta}$ and $\boldsymbol{\chi}_j$:

$$\boldsymbol{\zeta} = \left[\frac{1}{\sigma_z}\sum_{j=1}^{N}\mathbf{x}_j\right] - \frac{\bar{z}}{\sigma_z} = \left[\frac{1}{N^{1/2}\sigma_x}\sum_{j=1}^{N}\mathbf{x}_j\right] - \frac{N\bar{x}}{N^{1/2}\sigma_x}$$

$$= \frac{1}{N^{1/2}}\left[\sum_{j=1}^{N}\frac{\mathbf{x}_j - x}{\sigma_x}\right] = \frac{1}{N^{1/2}}\sum_{j=1}^{N}\boldsymbol{\chi}_j. \tag{3.77}$$

The characteristic function for $\boldsymbol{\zeta}$ is

$$M_\zeta(\lambda) = \langle \exp(-2\pi i\lambda\boldsymbol{\zeta})\rangle$$

$$= \left\langle \exp\left(-2\pi i\lambda N^{-1/2}\sum_{j=1}^{N}\boldsymbol{\chi}_j\right)\right\rangle. \tag{3.78}$$

However, because the $\boldsymbol{\chi}_j$ are independent and have the same statistics, this equation becomes

$$M_\zeta(\lambda) = \langle \exp(-2\pi i\lambda\boldsymbol{\chi}N^{-1/2})\rangle^N = [M_\chi(\lambda N^{-1/2})]^N, \tag{3.79}$$

where again the subscript j on $\boldsymbol{\chi}_j$ is superfluous within an expectation value.

If $M_\chi(\lambda N^{-1/2})$ is sufficiently well behaved, it can be represented as a Taylor series:

$$M_\chi(\lambda N^{-1/2}) = \left\langle \sum_{n=0}^{\infty}\left[-2\pi i\lambda\boldsymbol{\chi}N^{-1/2}\right]^n \Big/ n! \right\rangle = 1 - \frac{2\pi^2\lambda^2}{N} + \cdots, \tag{3.80}$$

where the term of order $\lambda N^{-1/2}$ vanishes since $\langle\boldsymbol{\chi}\rangle = 0$. The next term in the series varies as $N^{-3/2}$ if $\langle\boldsymbol{\chi}^3\rangle \neq 0$ and as N^{-2} if $\langle\boldsymbol{\chi}^3\rangle = 0$. As N approaches infinity, these higher-order terms become negligible. The term of order $1/N$ is, however, still important, as we can see by taking the logarithm of (3.79):

$$\ln M_\zeta(\lambda) = N\ln M_\chi(\lambda N^{-1/2})$$

$$= N\ln[1 - (2\pi^2\lambda^2/N)] \xrightarrow[N\to\infty]{} (-2\pi^2\lambda^2), \tag{3.81}$$

since

$$\lim_{\varepsilon\to 0}\ln(1+\varepsilon) = \varepsilon. \tag{3.82}$$

Therefore we have

$$M_\zeta(\lambda) = \exp(-2\pi^2\lambda^2) \tag{3.83}$$

and an inverse Fourier transform [see Eqs. (B.37) and (B.12)] yields

$$\mathrm{pr}_\zeta(\zeta) = \mathscr{F}_1^{-1}\{M_\zeta(\lambda)\} = (2\pi)^{-1/2}\exp(-\zeta^2/2). \tag{3.84}$$

Comparing this equation to the standard form for the Gaussian distribution, (3.58), we see that ζ is indeed a normal random variable with zero mean and unit variance as expected. In terms of the original variable z,

$$\mathrm{pr}_z(z) = \frac{d\zeta}{dz}\,\mathrm{pr}_\zeta(\zeta) = (2\pi\sigma_z^2)^{-1/2}\exp\left(\frac{-(z-\bar{z})}{2\sigma_z^2}\right), \tag{3.85}$$

where $\bar{z}$ and σ_z are given by (3.76).

3.3 POISSON RANDOM PROCESSES

In this section we introduce a type of noise that has its origin in the discrete, quantum nature of electromagnetic radiation. A radiation detector does not continuously absorb energy from the radiation field, but rather does so in increments of $h\nu$, where ν is the frequency of the radiation and h is Planck's constant. Therefore, the output of the detector cannot be a smooth, unvarying measure of the intensity of the field, but must exhibit fluctuations known variously as *quantum noise*, *photon noise*, or *Poisson noise*. Much the same problem arises in electronics where currents suffer fluctuations, usually called *shot noise*, due to the discrete nature of electrons.

Quantum noise plays a crucial role in radiographic imaging systems, primarily because $h\nu$ is so large for x rays and gamma rays. Coupled with the necessity of limiting the radiation dose delivered to the patient, the large energy per quantum means that the final image will consist of a relatively small number of detected quanta. As an extreme example, a nuclear medicine image of the heart—one frame of a dynamic cardiac study—usually consists of less than 5000 detected gamma rays. A more typical static nuclear image would consist of 500,000 detected quanta, while a diagnostic x-ray film used with a fluorescent screen would be exposed to a usable density with 10^7 x-ray photons/cm^2. For comparison, a typical photographic film exposed to light requires 10^{11}–10^{12} optical photons/cm^2 for a useful picture. With so few detected quanta, it is not surprising that quantum noise is usually the dominant form of noise in radiographic images. Accordingly, we shall give considerable attention to quantum noise in this book; the necessary mathematical background for its analysis is presented in this section.

3.3.1 Derivation of the Probability Law

The derivation given in this section closely parallels the discussion of electronic shot noise by Davenport and Root (1958).

Consider once again a detector viewing a radioisotope gamma-ray source. The isotope is assumed to have a very long half-life, and the source-detector geometry is held fixed so that stationary statistics are applicable. (This restriction will be removed in Section 3.3.4.) We also assume for simplicity that the detector is ideal and responds to every photon that impinges on it. However, as we shall show in Section 3.4.4, this assumption has no essential effect on the results.

Our immediate goal is to calculate the probability that exactly K photons are detected in an observation time T. Three physically reasonable assumptions about this probability, which we denote by $\Pr(K \text{ in } T)$, will be made:

(a) The number of photons detected in the interval $(0, T)$ is statistically independent of the number detected in any other nonoverlapping interval.

(b) The probability of detecting one photon in a vanishingly small time interval ΔT is directly proportional to ΔT, i.e.,

$$\lim_{\Delta T \to 0} \Pr(1 \text{ in } \Delta T) = a\,\Delta T, \tag{3.86}$$

where a is a constant.

(c) The probability of more than one photon being detected in ΔT is zero, so that

$$\lim_{\Delta T \to 0} \{\Pr(0 \text{ in } \Delta T) + \Pr(1 \text{ in } \Delta T)\} = 1. \tag{3.87}$$

Now let us determine the probability that *no* photons are detected in an interval of length $T + \Delta T$. The only way this can occur is if no photons are detected in the interval $(0, T)$ *and* no photons are detected in the adjacent interval $(T, T + \Delta T)$. By assumption (a), these two intervals are independent, so that

$$\Pr(0 \text{ in } T + \Delta T) = \Pr(0 \text{ in } T)\Pr(0 \text{ in } \Delta T). \tag{3.88}$$

By substitution of (3.86) and (3.87) into (3.88), we find

$$\lim_{\Delta T \to 0}\left(\frac{\Pr(0 \text{ in } T + \Delta T) - \Pr(0 \text{ in } T)}{\Delta T}\right) = \frac{d}{dT}\{\Pr(0 \text{ in } T)\}$$
$$= -a\Pr(0 \text{ in } T). \tag{3.89}$$

This elementary differential equation, along with obvious boundary condition, $\Pr(0 \text{ in } 0) = 1$, has the solution

$$\Pr(0 \text{ in } T) = \exp(-aT). \tag{3.90}$$

Now consider the probability that K photons are detected in the interval $T + \Delta T$. If $\Delta T \to 0$, only zero or one of these K photons can be detected in ΔT by assumption (c). There are thus two mutually exclusive ways to find

exactly K photons in $T + \Delta T$, viz., either K in T and 0 in ΔT, or $K - 1$ in T and 1 in ΔT. Again the intervals are independent, so that we may write

$$\begin{aligned}\Pr(K \text{ in } T + \Delta T) &= \Pr(K-1 \text{ in } T \text{ and } 1 \text{ in } \Delta T) + \Pr(K \text{ in } T \text{ and } 0 \text{ in } \Delta T)\\ &= \Pr(K-1 \text{ in } T)\Pr(1 \text{ in } \Delta T) + \Pr(K \text{ in } T)\Pr(0 \text{ in } \Delta T)\\ &= \Pr(K-1 \text{ in } T)a\Delta T + \Pr(K \text{ in } T)[1 - a\Delta T], \quad (3.91)\end{aligned}$$

where (3.86) and (3.87) were invoked in the last step. Rearranging the terms in (3.91) and passing to the limit, we find

$$\begin{aligned}&\lim_{\Delta T \to 0}\left(\frac{\Pr(K \text{ in } T + \Delta T) - \Pr(K \text{ in } T)}{\Delta T}\right) + a\Pr(K \text{ in } T)\\ &\quad = \frac{d}{dT}\{\Pr(K \text{ in } T)\} + a\Pr(K \text{ in } T) = a\Pr(K-1 \text{ in } T). \quad (3.92)\end{aligned}$$

If $\Pr(K-1 \text{ in } T)$ is known, (3.92) is a linear, inhomogeneous differential equation for the unknown $\Pr(K \text{ in } T)$; it may thus be regarded as a recursion relation for $\Pr(K \text{ in } T)$. By means of the usual integrating factor, the solution of (3.92) is found to be

$$\Pr(K \text{ in } T) = a\exp[-aT]\int_0^T \exp[aT']\Pr(K-1 \text{ in } T')dT', \quad (3.93)$$

where the boundary condition $\Pr(K \text{ in } 0) = 0$ has been used.

To use this recursion relation, we start with $K = 1$. Since we have already determined in (3.90) that $\Pr(0 \text{ in } T) = \exp(-aT)$,

$$\begin{aligned}\Pr(1 \text{ in } T) &= a\exp(-aT)\int_0^T \exp(aT')\exp(-aT')dT'\\ &= aT\exp(-aT). \quad (3.94)\end{aligned}$$

Continuing this way, we have

$$\begin{aligned}\Pr(2 \text{ in } T) &= a\exp(-aT)\int_0^T \exp(aT')\exp(-aT')aT'\,dT'\\ &= (a^2T^2/2)\exp(-aT), \quad (3.95)\\ \Pr(3 \text{ in } T) &= a\exp(-aT)\int_0^T \exp(aT')\exp(-aT')(a^2T^2/2)dT'\\ &= (a^3T^3/3\cdot 2)\exp(-aT), \quad (3.96)\end{aligned}$$

and so on.

The general rule is

$$\Pr(K \text{ in } T) = (a^KT^K/K!)\exp(-aT). \quad (3.97)$$

This result is the *Poisson probability law* or, loosely, the *Poisson distribution*.

3.3.2 Properties of Poisson Random Variables

Let us inquire more closely into the nature of the constant a that was introduced in (3.86) above. Since $a\Delta T$ is the probability that one photon is detected in ΔT, and the probability of more than one being detected vanishes as $\Delta T \to 0$, the quantity $a\Delta T$ may also be interpreted as the *mean* number of photons detected in ΔT. So, at least for the small interval ΔT, a is the mean number of detected photons *per unit time*. Furthermore, since we assumed stationary statistics at the outset, the mean number per unit time must be a constant and aT must be the mean number detected in T, i.e.,

$$\langle \mathbf{K} \rangle = aT. \quad (3.98)$$

This important result can also be verified by directly calculating $\langle \mathbf{K} \rangle$ from $\Pr(K \text{ in } T)$:

$$\begin{aligned}\langle \mathbf{K} \rangle &= \sum_{k=0}^{\infty} K\Pr(K \text{ in } T)\\ &= \exp(-aT)\sum_{k=0}^{\infty} K\left(\frac{a^KT^K}{K!}\right). \quad (3.99)\end{aligned}$$

To evaluate this sum, note first that the $K = 0$ term vanishes since $0! = 1$. This permits us to run the sum from $K = 1$ to ∞ rather than 0 to ∞. Next, define $L = K - 1$ and observe that $K/K! = 1/(K-1)! = 1/L!$. Therefore

$$\begin{aligned}\langle \mathbf{K} \rangle &= \exp(-aT)\sum_{L=0}^{\infty}\frac{a^{L+1}T^{L+1}}{L!}\\ &= \exp(-aT)aT\sum_{L=0}^{\infty}\frac{a^LT^L}{L!}. \quad (3.100)\end{aligned}$$

But this latter sum is merely $\exp(+aT)$, so that $\langle \mathbf{K} \rangle = aT$ as expected. Henceforth we shall write $\bar{K}$ for aT. Furthermore, since the time interval T no longer appears separately, we shall simplify the notation by writing $\Pr(K)$ for $\Pr(K \text{ in } T)$. Thus,

$$\Pr(K) = \exp(-\bar{K})\bar{K}^K/K!. \quad (3.101)$$

The procedure used to find $\langle \mathbf{K} \rangle$ can also be used to find the variance of K. We have

$$\begin{aligned}\sigma_K^2 &= \langle(\mathbf{K} - \bar{K})^2\rangle = \langle \mathbf{K}^2 \rangle - \bar{K}^2\\ &= \left[\exp(-\bar{K})\sum_{K=0}^{\infty}\left(\frac{\bar{K}^K}{K!}\right)K^2\right] - \bar{K}^2. \quad (3.102)\end{aligned}$$

Again, the $K = 0$ term vanishes so that

$$\sigma_K^2 = \left[\exp(-\bar{K}) \sum_{K=1}^{\infty} \left(\frac{\bar{K}^K}{(K-1)!}\right) K\right] - \bar{K}^2. \tag{3.103}$$

With $L = K - 1$, we may rewrite the sum

$$\begin{aligned}\sigma_K^2 &= \left[\exp(-\bar{K}) \sum_{L=0}^{\infty} \left(\frac{\bar{K}^{L+1}}{L!}\right)(L+1)\right] - \bar{K}^2 \\ &= \exp(-\bar{K})\left[\bar{K} \sum_{L=0}^{\infty} \left(\frac{\bar{K}^L}{L!}\right) L + \bar{K} \sum_{L=0}^{\infty} \left(\frac{\bar{K}^L}{L!}\right)\right] - \bar{K}^2. \end{aligned} \tag{3.104}$$

The first sum is readily evaluated since it is precisely the one that occurred in finding $\langle \mathbf{K} \rangle$. Therefore we already know, by comparison with (3.99), that

$$\sum_{L=0}^{\infty} \frac{L\bar{K}^L}{L!} = \bar{K} \exp(+\bar{K}). \tag{3.105}$$

The second sum in (3.104) is just $\exp(+\bar{K})$, so we have finally

$$\sigma_K^2 = \bar{K}. \tag{3.106}$$

This is the main distinguishing feature of Poisson random variables: *the variance always equals the mean.*

A simple corollary to this result is that the ratio of the mean $\bar{K}$ to the standard deviation σ_K, often called the signal-to-noise ratio or SNR, is given by

$$\text{SNR} = \bar{K}/\sigma_K = \bar{K}^{1/2}. \tag{3.107}$$

The rms accuracy of a counting experiment, expressed as a fraction of the mean count, is just the reciprocal of the SNR and hence the reciprocal of the square root of the mean number of counts.

Since $\bar{K}$ is the only parameter that appears in $\Pr(K)$, we should be able to state all moments of K in terms of $\bar{K}$. This is indeed possible, although we shall not derive it here. The result is given by Metz (1969) in the form of a recursion relation:

$$\langle \mathbf{K}^m \rangle = \bar{K} \sum_{l=0}^{m-1} \binom{m-1}{l} \langle \mathbf{K}^l \rangle, \tag{3.108}$$

where $\binom{m-1}{l}$ is the binomial coefficient defined by

$$\binom{m}{l} = \frac{m!}{l!(m-l)!}. \tag{3.109}$$

The first few moments are

$$\begin{aligned}\langle \mathbf{K}^2 \rangle &= \bar{K}^2 + \bar{K}, \\ \langle \mathbf{K}^3 \rangle &= \bar{K}^3 + 3\bar{K}^2 + \bar{K}, \\ \langle \mathbf{K}^4 \rangle &= \bar{K}^4 + 6\bar{K}^3 + 7\bar{K}^2 + \bar{K}. \end{aligned} \tag{3.110}$$

The first few central moments are

$$\begin{aligned}\langle (\mathbf{K} - \bar{K})^2 \rangle &= \bar{K}, \\ \langle (\mathbf{K} - \bar{K})^3 \rangle &= \bar{K}, \\ \langle (\mathbf{K} - \bar{K})^4 \rangle &= 3\bar{K}^2 + \bar{K}, \\ \langle (\mathbf{K} - \bar{K})^5 \rangle &= 10\bar{K}^2 + \bar{K}, \\ \langle (\mathbf{K} - \bar{K})^6 \rangle &= 15\bar{K}^3 + 5\bar{K}^2 + \bar{K}. \end{aligned} \tag{3.111}$$

Of course, all of the moments can be derived from the characteristic function, given by

$$\begin{aligned} M_K(\lambda) &= \langle \exp(-2\pi i \lambda \mathbf{K}) \rangle \\ &= \exp(-\bar{K}) \sum_{K=0}^{\infty} \left(\frac{\bar{K}^K}{K!}\right) \exp(-2\pi i \lambda K) \\ &= \exp(-\bar{K}) \sum_{K=0}^{\infty} [\bar{K} \exp(-2\pi i \lambda)]^K / K! \\ &= \exp\{-\bar{K}[1 - \exp(-2\pi i \lambda)]\}. \end{aligned} \tag{3.112}$$

This rather bizarre expression for $M_K(\lambda)$ makes it easy to determine the behavior of a Poisson distribution for large values of $\bar{K}$. The first point to be noted is that $M_K(\lambda)$ is a periodic function of λ, which is a consequence of $\mathbf{K}$ being a discrete random variable. Thus we can expand $M_K(\lambda)$ in a Fourier series of the form

$$M_K(\lambda) = \sum_{K=0}^{\infty} \Pr(K) \exp(-2\pi i K \lambda). \tag{3.113}$$

To see that the Fourier coefficients are just the $\Pr(K)$, we can take a Fourier *transform* of (3.113) and produce a probability density function in the form of a sum of delta functions as in (3.21). The $\Pr(K)$ are given by the usual formula for Fourier coefficients (B.83), which involves an integral over the basic period $-\frac{1}{2} < \lambda < \frac{1}{2}$. However, if $\bar{K}$ is large, $M_K(\lambda)$ will be small unless $1 - \exp(-2\pi i \lambda)$ is small. Within the basic period of λ, this condition implies that $|\lambda| \ll 1$, and immediately suggests an expansion of $\exp(-2\pi i \lambda)$ in powers

of λ. This yields, with (B.83),

$$\begin{aligned}\Pr(K) &= \int_{-1/2}^{1/2} M_K(\lambda)\exp(2\pi i K\lambda)\,d\lambda \\ &= \int_{-1/2}^{1/2} \exp[-\bar{K}(2\pi i\lambda + 2\pi^2\lambda^2 + \cdots)]\exp(2\pi i K\lambda)\,d\lambda \\ &\approx \int_{-\infty}^{\infty} \exp[-\bar{K}(2\pi i\lambda + 2\pi^2\lambda^2)]\exp(2\pi i K\lambda)\,d\lambda \\ &= (2\pi\bar{K})^{-1/2}\exp[-(K-\bar{K})^2/2\bar{K}], \end{aligned} \tag{3.114}$$

where the last step follows from (B.13), (B.14), and (B.37). Thus $\Pr(K)$ approaches a Gaussian for large $\bar{K}$, a result that could also have been anticipated from the central-limit theorem.

A graphical comparison between the exact Poisson distribution, (3.101), and the Gaussian approximation, (3.114), is given in Fig. 3.4. As a rule of

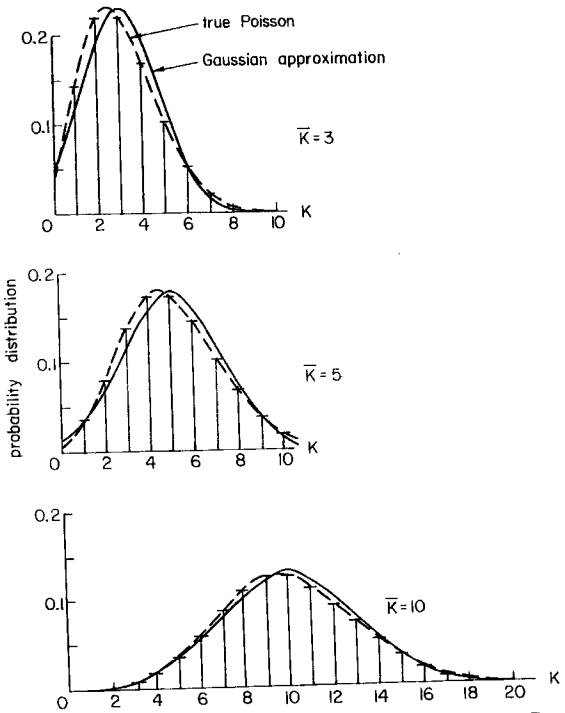

Fig. 3.4 A comparison between the exact Poisson distribution and the Gaussian approximation to it for $\bar{K} = 3$, 5, and 10. The Poisson distribution is defined only for integer values of K, indicated by the vertical bars, and the dotted line is simply a smooth curve through the points.

thumb for numerical calculations, the Gaussian form is an excellent approximation for $\bar{K} \geqslant 10$.

3.3.3 Poisson Impulses and Statistical Averages

The quantity $\Pr(K)$ does not fully describe the statistics of our hypothetical counting experiment. In many cases, we also need to know the statistics of the arrival times of individual photons. If the jth photon out of the group of K detected photons arrives at the detector at time $\mathbf{t}_j$, then $\mathbf{t}_j$ is a continuous random variable of domain $(0, T)$ and probability density function $\mathrm{pr}_t(t_j)$. We have chosen the subscript t rather than j or t_j for $\mathrm{pr}_t(t_j)$ since there is nothing special about the jth photon. The index j merely labels the various photons and says nothing about the order of arrival. The photon for which $j = 19$ could well arrive before the one with $j = 2$. All photons must have the same probability of arrival during any specified finite time interval. Indeed, it follows from assumption (a) of Section 3.3.1 and the assumption of stationary statistics that $\mathrm{pr}_t(t_j)$ is a constant over the interval $(0, T)$. For proper normalization, this constant must have the value T^{-1}, i.e.,

$$\mathrm{pr}_t(t_j) = \begin{cases} T^{-1} & \text{if } 0 < t_j < T \\ 0 & \text{otherwise.} \end{cases} \tag{3.115}$$

A more rigorous proof of this equation can be given (see Davenport and Root, 1958; Papoulis, 1965), but since it is so appealing on physical grounds, we shall regard it as an article of faith. Equation (3.115) is at least as "obvious" as assumption (a) from which it can be derived.

The arrival of photons at the detector can thus be described in terms of a random process $\mathbf{z}(t)$ given by

$$\mathbf{z}(t) = \sum_{j=1}^{\mathbf{K}} \delta(t - \mathbf{t}_j). \tag{3.116}$$

If each arriving photon produces a voltage pulse of the form $V_0(t - \mathbf{t}_j)$ at the output of the detector, then the overall voltage from the detector is the random process $\mathbf{V}(t)$ given by

$$\mathbf{V}(t) = \sum_{j=1}^{\mathbf{K}} V_0(t - \mathbf{t}_j) = V_0(t) * \mathbf{z}(t). \tag{3.117}$$

The process $\mathbf{z}(t)$, often called a sequence of *Poisson impulses*, is thus intimately related to the observed random voltage at the detector output. The statistics of $\mathbf{z}(t)$ will therefore be examined in some detail.

First, let us calculate the mean value of $\mathbf{z}(t)$. Note that there are $\mathbf{K} + 1$ random variables in this problem—the $\mathbf{K}$ arrival times $\mathbf{t}_j$ and $\mathbf{K}$ itself. All

of these random variables are statistically independent, and we have already determined $\Pr(K)$ and $\mathrm{pr}_t(t_j)$. Therefore we can write

$$\langle \mathbf{z}(t) \rangle = \sum_{K=0}^{\infty} \Pr(K) \int_0^T \mathrm{pr}_t(t_1)\,dt_1 \int_0^T \mathrm{pr}_t(t_2)\,dt_2 \cdots \int_0^T \mathrm{pr}_t(t_K)\,dt_K \sum_{j=1}^{K} \delta(t - t_j). \tag{3.118}$$

Now pick any one term in the sum over j, say the one for which $j = 17$. All of the integrals *except* the one over t_{17} have the form $\int_0^T \mathrm{pr}_t(t_j)\,dt_j$; from (3.115) each of these integrals is unity. The remaining integral is

$$\int_0^T \mathrm{pr}_t(t_{17})\,\delta(t - t_{17})\,dt_{17} = \mathrm{pr}_t(t) = T^{-1} \tag{3.119}$$

assuming that $0 < t < T$.

Of course the choice $j = 17$ was quite arbitrary, and it is reassuring that the number 17 does not appear in the result. Each of the K terms in the sum over j gives the same result, so that (3.118) becomes

$$\langle \mathbf{z}(t) \rangle = T^{-1} \sum_{K=0}^{\infty} K \Pr(K) = \frac{\bar{K}}{T}. \tag{3.120}$$

The expectation value of $\mathbf{z}(t)$ is thus the mean number of detected photons per unit time, which is the constant a from Section 3.3.1.

Now we turn to a slightly more complicated problem and calculate the autocorrelation function for $\mathbf{z}(t)$. By analogy to (3.118), we have

$$\begin{aligned} R_z(\tau) &= \left\langle \sum_{j=1}^{\mathbf{K}} \delta(t - \mathbf{t}_j) \sum_{k=1}^{\mathbf{K}} \delta(t + \tau - \mathbf{t}_k) \right\rangle \\ &= \sum_{K=0}^{\infty} \Pr(K) \int_0^T \mathrm{pr}_t(t_1)\,dt_1 \int_0^T \mathrm{pr}_t(t_2)\,dt_2 \cdots \\ &\quad \times \int_0^T \mathrm{pr}_t(t_K)\,dt_K \sum_{j=1}^{K} \sum_{k=1}^{K} \delta(t - t_j)\,\delta(t + \tau - t_k). \end{aligned} \tag{3.121}$$

There are a total of K^2 terms in the double sum over j and k. Of these, K terms have $j = k$, while the remaining $K^2 - K$ terms have $j \neq k$. Consider first the case where $j = k$. As in the calculation of $\langle \mathbf{z}(t) \rangle$, we pick out a particular term from this group of K terms, for example the one for which $j = k = 17$. Again all integrals except the one over t_{17} reduce to unity, while the one over t_{17} has the form

$$\int_0^T \mathrm{pr}_t(t_{17})\,\delta(t - t_{17})\,\delta(t + \tau - t_{17})\,dt_{17}. \tag{3.122}$$

Integrals with two delta functions in the argument present no difficulty if one of them is replaced with one of the various limiting forms for delta functions discussed in Appendix A [see (A.7)–(A.11)]. This delta function may then be treated as an ordinary function, and the sifting property (A.1) may be invoked for the other delta function in the integrand. In (3.122), by this argument, the delta function $\delta(t - t_{17})$ sifts out values of t_{17} for which $t - t_{17} = 0$. The expression (3.122) then becomes

$$\delta(\tau)\,\mathrm{pr}_t(t) = T^{-1}\,\delta(\tau). \tag{3.123}$$

Once again, the number 17 does not appear in the result, and all K terms have the same value. The contribution to $R_z(\tau)$ from terms in which $j = k$ is therefore

$$[R_z(\tau)]_{j=k} = T^{-1}\,\delta(\tau) \sum_{K=0}^{\infty} K \Pr(K) = \bar{K} T^{-1}\,\delta(\tau). \tag{3.124}$$

We now consider the $K^2 - K$ terms for which $j \neq k$. Consider a particular pair of values for j and k, e.g., $j = 17$ and $k = 9$. Now all integrals except the two over t_{17} and t_9 reduce to unity, while these two factor as

$$\int_0^T \mathrm{pr}_t(t_{17})\,\delta(t - t_{17})\,dt_{17} \int_0^T \mathrm{pr}_t(t_9)\,\delta(t + \tau - t_9) = \mathrm{pr}_t(t)\,\mathrm{pr}_t(t + \tau) = \frac{1}{T^2}, \tag{3.125}$$

where the last step requires both that $0 < t < T$ and that $0 < (t + \tau) < T$.

Neither 9 nor 17 appears in (3.125), so all $K^2 - K$ terms must yield the same result. The contribution to $R_z(\tau)$ from terms in which $j \neq k$ is thus

$$[R_z(\tau)]_{j \neq k} = T^{-2} \sum_{K=0}^{\infty} (K^2 - K) \Pr(K) = T^{-2} \langle \mathbf{K}^2 - \mathbf{K} \rangle. \tag{3.126}$$

However, from (3.110), $\langle \mathbf{K}^2 - \mathbf{K} \rangle = \bar{K}^2$.

Our final result for $R_z(\tau)$ is

$$\begin{aligned} R_z(\tau) &= [R_z(\tau)]_{j=k} + [R_z(\tau)]_{j \neq k} \\ &= \bar{K} T^{-1}\,\delta(\tau) + \bar{K}^2 T^{-2} = a\,\delta(\tau) + a^2, \end{aligned} \tag{3.127}$$

where a is again $\bar{K}/T$, the mean number detected per unit time. Since only a, and not T separately, appears in $R_z(\tau)$, we may as well let $T \to \infty$ and have no further worry about whether t or $t + \tau$ lies in the interval $(0, T)$.

Equation (3.127) will be important to our understanding of noise in radiographic imaging, but not quite in the form given. Instead, we shall more

often be interested in Poisson impulses that are randomly distributed in *space* rather than in *time*. Consider, for example, the spatial random process $\mathbf{u}(\mathbf{r})$, given by

$$\mathbf{u}(\mathbf{r}) = \sum_{j=1}^{K} \delta(\mathbf{r} - \mathbf{r}_j), \tag{3.128}$$

where $\mathbf{r}$ and $\mathbf{r}_j$ are both two-dimensional vectors but only the latter is a random variable. The process $\mathbf{u}(\mathbf{r})$ could represent the distribution of recorded photons on the crystal of an Anger camera or on a piece of x-ray film. If the detector area is A_d and the photons are uniformly sprinkled over this area, the probability density function for each of the random variables $\mathbf{r}_j$ is given by

$$\mathrm{pr}_{\mathbf{r}}(\mathbf{r}_j) = A_d^{-1} \tag{3.129}$$

if $\mathbf{r}_j$ lies in A_d. Furthermore, $\mathbf{K}$ is a Poisson random variable since the fact that the detector is capable of determining the position of each photon, as well as merely recording its presence, has no effect on the total count. The derivation of $\Pr(K)$ given in Section (3.3.1) still holds.

There is thus a complete formal equivalence between $\mathbf{z}(t)$ and $\mathbf{u}(\mathbf{r})$, and we do not need to repeat the calculation. We can write at once that

$$\langle \mathbf{u}(\mathbf{r}) \rangle = \bar{K}/A_d \tag{3.130}$$

and

$$R_u(\mathbf{L}) = \langle \mathbf{u}(\mathbf{r})\mathbf{u}(\mathbf{r} + \mathbf{L}) \rangle = (\bar{K}/A_d)\,\delta(\mathbf{L}) + (\bar{K}/A_d)^2, \tag{3.131}$$

where $\mathbf{L}$ is a two-dimensional position vector and we assume that both $\mathbf{r}$ and $\mathbf{r} + \mathbf{L}$ fall within A_d.

Equation (3.131) will serve as the starting point for our analysis of noise in radiographic imaging.

3.3.4 Nonstationary Poisson Processes

Our entire discussion of Poisson random processes to this point has assumed stationary statistics. In the case of temporal Poisson processes, the mean number of detected photons per unit time was assumed to be independent of time, the probability density for the arrival times was assumed to be constant, and the autocorrelation function for a sum of Poisson impulses was shown to depend only on the time difference τ and not on the actual time t. Similarly, in our brief discussion of spatial Poisson processes, we assumed that all locations $\mathbf{r}_j$ were equally likely and that the autocorrelation function for $\mathbf{u}(\mathbf{r})$ depended on the difference vector $\mathbf{L}$ and not on the actual position $\mathbf{r}$.

Now, in real physical problems, we can be almost guaranteed that stationary statistics are not exactly correct. In the temporal case, radioactive decay, variations in source-detector geometry, and simply turning off the apparatus at the end of the day all cause deviations from strict temporal stationarity.

Nonstationarity is even more important for spatial Poisson processes since a stationary spatial distribution of photons is not an image except in the most trivial sense. As the examples given in Chapter 1 show, it is the deviations from uniformity in a radiographic image that convey useful diagnostic information.

Fortunately, it is not difficult to generalize our discussion of Poisson processes to the nonstationary case. Let us first review the derivation of $\Pr(K)$ for the case where a, the mean arrival rate of photons, is a known (nonrandom) function of time. Since it is straightforward to retrace the arguments of Section 3.3.1, we simply present here the generalized equations without commentary:

$$(3.86) \rightarrow \lim_{\Delta T \to 0} \Pr(1 \text{ in } \Delta T) = a(T)\,\Delta T, \tag{3.132}$$

$$(3.89) \rightarrow \frac{d}{dT} \Pr(0 \text{ in } T) = -a(T) \Pr(0 \text{ in } T), \tag{3.133}$$

$$(3.90) \rightarrow \Pr(0 \text{ in } T) = \exp\left(-\int_0^T a(t)\,dt \right) \tag{3.134}$$

$$(3.91) \rightarrow \Pr(K \text{ in } T + \Delta T) = \Pr(K - 1 \text{ in } T)a(T)\,\Delta T + \Pr(K \text{ in } T)[1 - a(T)\,\Delta T], \tag{3.135}$$

$$(3.92) \rightarrow \frac{d}{dT} \Pr(K \text{ in } T) + a(T) \Pr(K \text{ in } T) = a(T) \Pr(K - 1 \text{ in } T), \tag{3.136}$$

$$(3.93) \rightarrow \Pr(K \text{ in } T) = \exp\left(-\int_0^T a(t)\,dt \right) \int_0^T a(T') \times \exp\left(\int_0^{T'} a(T'')\,dT'' \right) \Pr(K - 1 \text{ in } T')\,dT', \tag{3.137}$$

$$(3.94) \rightarrow \Pr(1 \text{ in } T) = \exp\left(-\int_0^T a(t)\,dt \right) \int_0^T a(T')\,dT', \tag{3.138}$$

$$(3.97) \rightarrow \Pr(K \text{ in } T) = \left\{ \left(\int_0^T a(T')\,dT' \right)^K \Big/ K! \right\} \exp\left(-\int_0^T a(t)\,dt \right). \tag{3.139}$$

In other words, our previous expression for $\Pr(K)$ is still valid so long as we take

$$\bar{K} = \int_0^T a(t)\,dt, \tag{3.140}$$

instead of just aT.

The discussion of Poisson impulses in Section 3.3.3 also generalizes readily. The point of departure is (3.115) which, in the nonstationary case, must become

$$\mathrm{pr}_t(t_j) = a(t_j) \Big/ \int_0^T a(t')\,dt' = a(t_j)/\bar{K}, \tag{3.141}$$

provided $0 < t_j < T$. Equation (3.141) merely states that the probability of any one photon arriving in some specific time interval is proportional to the mean number arriving in that interval. The main effect of this modification is to replace T^{-1} by $a(t)/\bar{K}$ at most places in Section 3.3.3. For example, (3.120) becomes

$$\langle \mathbf{z}(t) \rangle = a(t). \tag{3.142}$$

A little care must be exercised in finding the autocorrelation function for $\mathbf{z}(t)$ in the nonstationary case. The basic definition is [cf. (3.121)]

$$R_z(t, t+\tau) = \left\langle \sum_{j=1}^{\mathbf{K}} \delta(t - \mathbf{t}_j) \sum_{k=1}^{\mathbf{K}} \delta(t + \tau - \mathbf{t}_k) \right\rangle. \tag{3.143}$$

Most of the derivation is the same except that (3.125) becomes

$$\mathrm{pr}_t(t)\,\mathrm{pr}_t(t+\tau) = a(t)a(t+\tau)/\bar{K}^2. \tag{3.144}$$

The final result is

$$R_z(t, t+\tau) = a(t)\delta(\tau) + a(t)a(t+\tau). \tag{3.145}$$

The most obvious difference between $R_z(t, t+\tau)$ in (3.145) and $R_z(\tau)$ in (3.127) is in the second term. Yet even this difference is rather insignificant as we can see by considering the zero-mean random process $\Delta\mathbf{z}(t)$ defined by

$$\Delta\mathbf{z}(t) = \mathbf{z}(t) - \langle \mathbf{z}(t) \rangle. \tag{3.146}$$

In the nonstationary case, with $\langle \mathbf{z}(t) \rangle = a(t)$, we see that

$$\begin{aligned} R_{\Delta z}(t, t+\tau) &= \langle \Delta\mathbf{z}(t)\,\Delta\mathbf{z}(t+\tau) \rangle \\ &= R_z(t, t+\tau) - \langle \mathbf{z}(t) \rangle \langle \mathbf{z}(t+\tau) \rangle \\ &= a(t)\,\delta(\tau). \end{aligned} \tag{3.147}$$

The corresponding result for the stationary case is quite similar:

$$R_{\Delta z}(\tau) = a\,\delta(\tau). \tag{3.148}$$

The treatment of spatial Poisson impulses is identical. Equation (3.129) becomes

$$\mathrm{pr}_{\mathbf{r}}(\mathbf{r}_j) = h(\mathbf{r}_j) \Big/ \left(\int_{\mathrm{det}} h(\mathbf{r})\,d^2r \right), \tag{3.149}$$

where $h(\mathbf{r})$ is the mean number of detected photons per unit area, and the integral runs over the area of the detector. The random process $\mathbf{u}(\mathbf{r})$ then has a mean of

$$\langle \mathbf{u}(\mathbf{r}) \rangle = h(\mathbf{r}), \tag{3.150}$$

and an autocorrelation function of

$$R_u(\mathbf{r}, \mathbf{r}+\mathbf{L}) = h(\mathbf{r})\,\delta(\mathbf{L}) + h(\mathbf{r})h(\mathbf{r}+\mathbf{L}). \tag{3.151}$$

For the zero-mean process $\Delta\mathbf{u}(\mathbf{r}) = \mathbf{u}(\mathbf{r}) - \langle \mathbf{u}(\mathbf{r}) \rangle$, we have

$$R_{\Delta u}(\mathbf{r}, \mathbf{r}+\mathbf{L}) = h(\mathbf{r})\,\delta(\mathbf{L}). \tag{3.152}$$

3.4 THE BERNOULLI DISTRIBUTION

The Bernoulli or binomial distribution arises in binary selection processes (Fig. 3.5) in which there are only two possible outcomes of a basic experiment. For example, a coin is flipped n times. The only possible outcomes of each flip are heads, which has probability p, and tails, which has probability $q = 1 - p$. The number of heads in n trials is the random variable $\mathbf{k}$, with probability $\Pr(k|n)$, which is written as a conditional probability in anticipation of n becoming a random variable. For now, however, n is a fixed number.

As a second example, consider a detector on which exactly n photons fall during an observation time. Each photon is either detected or it is not; again the selection process is binary. If the probability of detection is p, then the probability that k out of the n photons are detected is the same $\Pr(k|n)$ as in the coin flipping case.

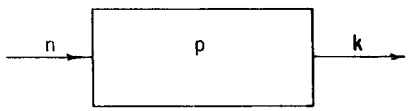

Fig. 3.5 Block diagram of a system described by a Bernoulli distribution. Exactly n items are presented at the input, with each of them being either accepted (probability p) or rejected (probability $q = 1 - p$).

3.4.1 Binomial Probability

Suppose we are given exactly n identical items and accept or reject each randomly. We assume that each item has a probability of selection of p, and that each selection stage is independent of all other stages. One way in which exactly k items could be selected would be to select the first k and reject the remaining $n-k$ (where $n \geqslant k$). The probability of this particular sequence occurring is $p^k(1-p)^{n-k}$. However, this is by no means the only way in which k items could be selected. For example, the first item could be rejected, the next k selected, and the final $n-k-1$ rejected. This sequence would also have probability $p^k(1-p)^{n-k}$. It is well known that the total number of ways of selecting k items from a group of n items is the binomial coefficient $\binom{n}{k}$ defined in (3.109). Each of these ways of selecting k items is mutually exclusive of all other ways. Therefore

$$\Pr(k|n) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{if} \quad n > k. \tag{3.153}$$

Of course, k items cannot be selected if there are less than k to choose from, so $\Pr(k|n) = 0$ if $n < k$. This expression for $\Pr(k|n)$ is called the Bernoulli or binomial distribution.

3.4.2 First and Second Moments

To calculate the moments of $\mathbf{k}$, we shall first determine the characteristic function (Papoulis, 1965). From (3.37)

$$\begin{aligned} M_k(\lambda) &= \langle \exp(-2\pi i \lambda \mathbf{k}) \rangle = \sum_{k=0}^{n} \binom{n}{k} \exp(-2\pi i \lambda k) p^k q^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} [p \exp(-2\pi i \lambda)]^k q^{n-k}, \end{aligned} \tag{3.154}$$

where $q = 1 - p$. The binomial theorem of algebra states that

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{3.155}$$

Therefore,

$$M_k(\lambda) = [p \exp(-2\pi i \lambda) + q]^n. \tag{3.156}$$

Application of (3.39) shows that

$$\bar{k} = \langle \mathbf{k} \rangle = (-2\pi i)^{-1} [dM_k(\lambda)/d\lambda]_{\lambda=0} = np(p+q)^{n-1} = np, \tag{3.157}$$

since $(p+q) = 1$. Similarly,

$$\begin{aligned} \langle \mathbf{k}^2 \rangle &= (-2\pi i)^{-2} [d^2 M_k(\lambda)/d\lambda^2]_{\lambda=0} = np + n(n-1)p^2 \\ &= npq + n^2 p^2. \end{aligned} \tag{3.158}$$

Therefore the variance of $\mathbf{k}$ is

$$\sigma_k^2 = \langle \mathbf{k}^2 \rangle - \langle \mathbf{k} \rangle^2 = npq = np(1-p). \tag{3.159}$$

3.4.3 Limiting Behavior

Consider the limit $n \to \infty$ and $p \to 0$ in such a way that np remains constant at $\bar{k}$. The characteristic function may be written, in terms of $\bar{k}$ and n,

$$M_k(\lambda) = \left(1 - \frac{\bar{k}}{n} + \frac{\bar{k}}{n} \exp(-2\pi i \lambda)\right)^n. \tag{3.160}$$

But it is well known that

$$\lim_{n \to \infty} (1 + \alpha/n)^n = e^{\alpha}. \tag{3.161}$$

This equation is, in fact, often regarded as a definition of e. Therefore

$$\lim_{n \to \infty} M_k(\lambda) = \exp\{-\bar{k}[1 - \exp(-2\pi i \lambda)]\}. \tag{3.162}$$

Comparison with (3.112) shows that (3.162) is just the characteristic function for the Poisson distribution. Thus the binomial distribution approaches the Poisson in this limit. That this result is reasonable can also be seen from (3.157) and (3.159):

$$\lim_{n \to \infty} [\sigma_k^2 / \langle \mathbf{k} \rangle] = \lim_{n \to \infty} q = \lim_{n \to \infty} (1 - \bar{k}/n) = 1. \tag{3.163}$$

Thus, in this limit, the variance equals the mean, which is the hallmark of the Poisson distribution.

3.4.4 Cascaded Random Processes—Poisson Source

The binomial distribution is applicable when $\mathbf{k}$ items are selected at random from a *fixed* number of possibilities n. In most physical problems, however, the number of possibilities should itself be a random variable $\mathbf{n}$. For example, $\mathbf{n}$ optical photons may strike a photomultiplier, with $\mathbf{k}$ of them producing photoelectrons (where $\mathbf{k} \leqslant \mathbf{n}$). Or $\mathbf{n}$ gamma-ray photons may be emitted by a source, with $\mathbf{k}$ of them producing a count in a distant detector. In these problems, the binomial distribution describes the conditional probability $\Pr(k|n)$ but not the overall probability $\Pr(k)$, which is

given by [cf. (3.9) and (3.11)]

$$\Pr(k) = \sum_{n=0}^{\infty} \Pr(k|n) \Pr(n). \tag{3.164}$$

We now consider the form of $\Pr(k)$ when $\Pr(n)$ is Poisson (Fried, 1965). Inserting (3.153) for $\Pr(k|n)$ and (3.101) for $\Pr(n)$ into (3.164), we find

$$\Pr(k) = \sum_{n=k}^{\infty} \exp(-\bar{n}) \frac{\bar{n}^n}{n!} \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}. \tag{3.165}$$

In spite of the formidable appearance of this equation, substantial simplifications are possible. Letting $m = n - k$, we have

$$\begin{aligned} \Pr(k) &= \frac{\exp(-\bar{n})p^k}{k!} \sum_{m=0}^{\infty} \frac{\bar{n}^{m+k}}{m!} (1-p)^m \\ &= \frac{\exp(-\bar{n})(p\bar{n})^k}{k!} \sum_{m=0}^{\infty} \frac{(\bar{n} - p\bar{n})^m}{m!}. \end{aligned} \tag{3.166}$$

The sum is just $\exp(\bar{n} - p\bar{n})$, so

$$\Pr(k) = \exp(-p\bar{n})(p\bar{n})^k/k!. \tag{3.167}$$

In other words, $\Pr(k)$ is simply a Poisson distribution with $\bar{k} = p\bar{n}$.

This result is applicable to many cascade situations. If a source emits $\mathbf{n}$ photons in random directions, where $\mathbf{n}$ is a Poisson random variable, then (3.167) gives the probability that k of them strike a detector if p is interpreted as the fractional solid angle $\Omega/4\pi$ subtended by the detector. Since $\Pr(k)$ is still Poisson, the same result can be applied again to find the number actually detected. In this case, p is the quantum efficiency η of the detector. Thus the number of detected photons $\mathbf{k}_d$ is also Poisson with mean $\bar{k}_d = \eta\bar{k} = (\eta\Omega/4\pi)\bar{n}$. As a practical matter, (3.167) allows us to use simple, deterministic arguments to calculate the mean value of the random variable at the end of the cascade, and then to be confident that the fluctuations about the mean are Poisson if the source is Poisson.

3.4.5 Cascaded Random Processes—Non-Poisson Source

We next consider the case where $\mathbf{n}$ does not obey Poisson statistics, but has some unspecified probability distribution $\Pr(n)$. However, we shall not calculate $\Pr(k)$, but only the mean and variance of $\mathbf{k}$.

The mean is given by

$$\langle \mathbf{k} \rangle = \sum_{k=0}^{\infty} k \Pr(k) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} k \Pr(k|n) \Pr(n). \tag{3.168}$$

The conditional probability is given by (3.153) and the conditional expectation value by (3.157). Therefore

$$\langle \mathbf{k} \rangle = \sum_{n=0}^{\infty} np \Pr(n) = \bar{n}p. \tag{3.169}$$

The variance of $\mathbf{k}$ is given by

$$\sigma_k^2 = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (k - \bar{n}p)^2 \Pr(k|n) \Pr(n). \tag{3.170}$$

According to (3.157) and (3.158), the conditional average over k has the effect of replacing k with np and k^2 with $npq + n^2p^2$, leaving

$$\sigma_k^2 = \sum_n [npq + n^2p^2 - 2\bar{n}np^2 + \bar{n}^2p^2] \Pr(n). \tag{3.171}$$

The remaining average over n replaces n with $\bar{n}$ and n^2 with its average $\langle \mathbf{n}^2 \rangle$, which we write

$$\langle \mathbf{n}^2 \rangle = \sum_n n^2 \Pr(n) = \sigma_n^2 + \bar{n}^2. \tag{3.172}$$

Some simple algebra then yields

$$\sigma_k^2 = \bar{n}pq + \sigma_n^2 p^2. \tag{3.173}$$

A more revealing form of this equation arises if we use $q = 1 - p$ and $\bar{k} = p\bar{n}$ to obtain

$$\sigma_k^2 - \bar{k} = p^2(\sigma_n^2 - \bar{n}). \tag{3.174}$$

If $\mathbf{n}$ were a Poisson random variable, we would have $\sigma_n^2 = \bar{n}$ and (3.174) would imply that $\sigma_k^2 = \bar{k}$, which is consistent with $\mathbf{k}$ also being a Poisson random variable. More generally, (3.174) shows that the *excess* noise (the variance minus the mean) scales as p^2 when $\mathbf{n}$ is subjected to a binomial selection with probability p. By contrast, if the noise were entirely Poisson, we would have $\sigma_k^2 = p\sigma_n^2$.

An algebraically equivalent form of (3.174) is

$$\frac{\sigma_k^2}{\bar{k}^2} = \frac{\sigma_n^2}{\bar{n}^2} + \frac{(1-p)p\bar{n}}{p^2\bar{n}^2} \tag{3.175}$$

or

$$\frac{1}{(\mathrm{SNR}_k)^2} = \frac{1}{(\mathrm{SNR}_n)^2} + \frac{1}{(\mathrm{SNR}_p)^2}, \tag{3.176}$$

where SNR_k and SNR_n are the ratio of the mean to the standard deviation for $\mathbf{k}$ and $\mathbf{n}$, respectively. SNR_p, on the other hand, is the *conditional* SNR

associated with the binomial process; i.e., it is the ratio of the mean of $\mathbf{k}$ to its standard deviation on the condition that the input to the process is exactly $\bar{n}$.

3.4.6 Cross Correlation of the Input and Output

In Section 3.4.7 we shall need to know the cross correlation $\langle \Delta\mathbf{n}\,\Delta\mathbf{k}\rangle$. The procedure for calculating it is the same as the one we just used. The required average is

$$\langle \Delta\mathbf{n}\,\Delta\mathbf{k}\rangle = \sum_n \sum_k (n-\bar{n})(k-\bar{n}p)\Pr(k|n)\Pr(n). \tag{3.177}$$

The conditional average over k simply replaces k with np, producing

$$\langle \Delta\mathbf{n}\,\Delta\mathbf{k}\rangle = \sum_n (n-\bar{n})(np-\bar{n}p)\Pr(n) = p\sum_n (n-\bar{n})^2\Pr(n) = p\sigma_n^2, \tag{3.178}$$

where the last step follows by the definition of the variance.

3.4.7 The Independent Nature of Poisson Random Variables

We shall repeatedly encounter situations in which two or more detectors are receiving radiation from a common source, and it will simplify the calculations if we can assume that the photon counts from different detectors are independent random variables. However, it is not obvious that this is a valid assumption. The source is fluctuating in the sense that the number of photons it emits in one counting interval is not the same as the number it emits in a different interval. The number emitted is the random variable $\mathbf{N}_0$, while the number detected by detector A is the random variable $\mathbf{N}_A$, and the number detected by B is $\mathbf{N}_B$. One might expect that when N_0 happened to be high due to a chance fluctuation, both N_A and N_B would also be high, and conversely if N_0 happened to be low. If this were the case, $\langle \Delta\mathbf{N}_A \Delta\mathbf{N}_B\rangle$ would be nonzero and $\mathbf{N}_A$ and $\mathbf{N}_B$ would not be independent.

Such a situation would be thoroughly inconsistent with the formalism developed in Sections 3.3.3 and 3.3.4. We considered there the Poisson random process $\mathbf{u}(\mathbf{r})$ [see (3.128)], in terms of which

$$\mathbf{N}_A = \int_A \mathbf{u}(\mathbf{r})\,d^2r, \tag{3.179}$$

$$\mathbf{N}_B = \int_B \mathbf{u}(\mathbf{r}')\,d^2r', \tag{3.180}$$

where the integrals run over the areas of the respective detectors. The cross correlation $\langle \Delta\mathbf{N}_A \Delta\mathbf{N}_B\rangle$ is given, with the aid of (3.152), by

$$\langle \Delta\mathbf{N}_A \Delta\mathbf{N}_B\rangle = \int_A d^2r \int_B d^2r'\, h(\mathbf{r})\,\delta(\mathbf{r}-\mathbf{r}') = 0, \tag{3.181}$$

since the argument of the delta function never vanishes if area A and area B do not overlap.

Thus, if $\mathbf{u}(\mathbf{r})$ is a valid description of the process, $\Delta\mathbf{N}_A$ and $\Delta\mathbf{N}_B$ must be uncorrelated. To see when this condition holds, let us reexamine the problem from the beginning without imposing any restrictions on the source. The discussion will rely heavily on results from Section 3.4.5.

The photons detected by detector A are a random (binomial) selection from the total number $\mathbf{N}_t = \mathbf{N}_A + \mathbf{N}_B$ that are detected by the two detectors. The probability that one of the *detected* photons is, in fact, detected by A is p_A, given by

$$p_A = \bar{N}_A/\bar{N}_t = \bar{N}_A/(\bar{N}_A + \bar{N}_B). \tag{3.182}$$

Similarly, the probability that one of the detected photons is detected by B is

$$p_B = \bar{N}_B/\bar{N}_t = 1 - p_A. \tag{3.183}$$

From (3.174), the statistics of $\mathbf{N}_A$ are related to those of $\mathbf{N}_t$ by

$$\sigma_{N_A}^2 - \bar{N}_A = p_A^2[\sigma_{N_t}^2 - \bar{N}_t]. \tag{3.184}$$

The cross correlation of $\Delta\mathbf{N}_A$ and $\Delta\mathbf{N}_t$ is given by (3.178) as

$$\langle \Delta\mathbf{N}_A\,\Delta\mathbf{N}_t\rangle = p_A\sigma_{N_t}^2. \tag{3.185}$$

Therefore

$$\begin{aligned}\langle \Delta\mathbf{N}_A\,\Delta\mathbf{N}_B\rangle &= \langle \Delta\mathbf{N}_A(\Delta\mathbf{N}_t - \Delta\mathbf{N}_A)\rangle = \langle \Delta\mathbf{N}_A\,\Delta\mathbf{N}_t\rangle - \sigma_{N_A}^2 \\ &= p_A\sigma_{N_t}^2 - \bar{N}_A - p_A^2[\sigma_{N_t}^2 - \bar{N}_t].\end{aligned} \tag{3.186}$$

But, by (3.182), $\bar{N}_A = p_A\bar{N}_t$, whence

$$\langle \Delta\mathbf{N}_A\,\Delta\mathbf{N}_B\rangle = p_A p_B(\sigma_{N_t}^2 - \bar{N}_t), \tag{3.187}$$

where we have also used $p_B = 1 - p_A$. This result shows that $\Delta\mathbf{N}_A$ and $\Delta\mathbf{N}_B$ are uncorrelated if $\mathbf{N}_t$ obeys Poisson statistics, for in that case $\sigma_{N_t}^2 - \bar{N}_t = 0$.

We can now go one step further and relate $\langle \Delta\mathbf{N}_A\,\Delta\mathbf{N}_B\rangle$ back to the properties of the source. Use of (3.174) shows that

$$\sigma_{N_t}^2 - \bar{N}_t = \left(\frac{\bar{N}_A + \bar{N}_B}{\bar{N}_0}\right)^2(\sigma_{N_0}^2 - \bar{N}_0), \tag{3.188}$$

where $\mathbf{N}_0$ is the number of photons emitted by the source. If we define overall detection probabilities f_A and f_B by

$$\begin{aligned} f_A &= \bar{N}_A/\bar{N}_0 = p_A(\bar{N}_A + \bar{N}_B)/\bar{N}_0, \\ f_B &= \bar{N}_B/\bar{N}_0 = p_B(\bar{N}_A + \bar{N}_B)/\bar{N}_0, \end{aligned} \tag{3.189}$$

then we have

$$\langle \Delta\mathbf{N}_A\,\Delta\mathbf{N}_B\rangle = f_A f_B(\sigma_{N_0}^2 - \bar{N}_0). \tag{3.190}$$

The important conclusion is that *Poisson sources cannot produce correlated counts in two different detectors.* Only the excess noise, beyond that associated with a Poisson source, can produce correlations. Poisson random variables are inherently uncorrelated (Hanbury Brown and Twiss, 1956).

3.4.8 A Note on the Applicability of Poisson Statistics

Since the Poisson distribution plays such a fundamental role in radiographic imaging, we shall reexamine the conditions under which it applies, making use of the discussion of Bernoulli processes just given. The assumptions used in the derivation of the Poisson distribution in Section 3.3.1 can be summarized by saying that we are dealing with *identical, independent, random* events.

Whether or not the events are sufficiently near to being identical depends on what measurement we are performing. In a photon-counting experiment, all that matters is that the event is recorded. We may select certain events on the basis of energy, spatial location, or other parameters, but the event is recorded in a binary fashion—either it occurred and met all criteria for recording or it did not. All recorded events are evidently identical. On the other hand, if we record high-energy photons by allowing them to expose photographic film, an additional degree of randomness is introduced. The amount of film blackening produced by each photon is itself a random variable, and the recorded events are no longer identical. In this case, the Poisson distribution cannot be rigorously justified, although it may well be a useful approximation.

The word *independent* in this context means that the number recorded in a time interval ΔT_1 is statistically independent of the number recorded in a different, nonoverlapping interval ΔT_2. To appreciate this restriction, let us construct a rather farfetched counterexample. Suppose we prepare a large number of samples of a radioactive material, and arrange for each sample to contain *exactly* 12 radioactive nuclei. We then observe each sample for two half-lives, so that the *mean* number of decays is $\frac{1}{2} \cdot 12 + \frac{1}{2} \cdot 6 = 9$. The histogram of the measured number of decays is a good approximation to the probability distribution for observing K decays in two half-lives, which we know to be a Bernoulli. It is easily seen that this distribution cannot be Poisson, because the maximum possible value for K is 12, while a Poisson distribution extends to $K = \infty$. Furthermore, the deviations from a Poisson must be substantial in this example because the difference between the mean (9) and the maximum (12) is just one standard deviation ($\sqrt{9} = 3$). The Poisson distribution thus predicts a fairly large probability of observing, say, $K = 13$, which we know is impossible.

The basic reason for the distribution in this example not being Poisson is that the number of counts in different time intervals is not independent. Suppose that all 12 of the radioactive nuclei happened to decay during the first half-life, an unlikely but not impossible occurrence. Then the number of decays in the second half-life is completely determined—it can only be zero. Similarly, whatever the number of decays in the first half-life, they determine the number of radioactive atoms left and therefore influence the number of decays in the second half-life.

Of course, all of this discussion is based on the premise that the various samples all contain exactly 12 radioactive nuclei. If we modify the experimental conditions somewhat (and make them more realistic) by assuming that the samples are drawn at random from a large supply of material, then the situation is quite different. If the number of radioactive nuclei in each sample is Poisson distributed with mean 12, then a calculation analogous to the one presented in Section 3.4.4 shows that the number of decays in two half-lives is also Poisson.

A more interesting question concerns the conditions that must be placed on the supply of material from which the samples are drawn in order for the number of radioactive nuclei per sample to be Poisson distributed. Suppose the supply contains a total of N_s nuclei, pN_s of which are radioactive and qN_s of which are nonradioactive. If we choose M nuclei at random, the probability that K of them will be radioactive is the Bernoulli distribution (3.153), which in the present notation is

$$\Pr(K) = \binom{M}{K} p^K q^{M-K}. \tag{3.191}$$

If, however, the supply N_s is very large but consists mainly of nonradioactive nuclei (p small), then this equation may be a good approximation to a Poisson. To be precise, consider the limit $N_s \to \infty$ and $p \to 0$. The number of nuclei in each sample M must also tend to ∞ if each sample is to have a nonzero number of radioactive nuclei. If $M \to \infty$ and $p \to 0$ in such a way that their product pM approaches a constant, then, from Section 3.4.3, $\Pr(K)$ approaches a Poisson of mean $\bar{K} = pM$.

From this result we can be virtually guaranteed that Poisson statistics are applicable in radioactive counting problems, because the supply is ultimately enormous. Whatever sample we have in our laboratory can be considered as a random selection from some larger supply which in turn is a selection from a still larger supply. The ultimate supply (the universe?) clearly satisfies the limiting condition for a Poisson distribution, while the cascaded selection process, by the result of Section 3.4.4, preserves the Poisson property for all smaller samples.

One final comment concerns the statistics of nonstationary counting problems. In Section 3.3.4 we showed that a Poisson distribution was obtained in the nonstationary case if the mean photon arrival rate $a(t)$ was a deterministic function of time. Examples where this is the case are easily constructed. For example, in radioactive decay problems we have $a(t) \propto \exp(-t/\tau)$. Also, if the source is modulated in some manner, $a(t)$ is under the

control of the experimenter. In these cases, the *ensemble* distribution of counts over any specified time interval must be Poisson, although the distribution observed in successive measurements on one member of the ensemble of apparatuses will, in general, not be Poisson.

The situation is very different if $a(t)$ is itself a random variable. Consider, for example, a photomultiplier detecting visible photons from a gas discharge lamp. The average intensity of the light source is quite random because of variations in current through the lamp and interference of radiation emitted independently by different atoms in the gas. The photocount distribution, even in an ensemble average sense, must then involve the distribution of the random variable $\mathbf{a}(t)$. The uncertainty in $\mathbf{a}(t)$ produces a further broadening of the photocount distribution over and above that predicted by Poisson statistics, and in general, the variance will exceed the mean. The Poisson distribution, where the variance equals the mean, normally represents the irreducible minimum spread in photon counting distributions, and occurs only when the mean arrival rate $a(t)$ is a deterministic function.*

3.5 FILTERING OF NOISY SIGNALS

As we showed in Chapter 2, imaging systems may be regarded as linear filters. Furthermore, even after an image is formed, it is often desirable to pass it through another linear filter. This second filter could be for image enhancement—for example, noise smoothing or deblurring—or it could be essential to the imaging process as in computed tomography and coded-aperture imaging. In any case, we cannot fully analyze radiographic images unless we know the effect of linear filters on noisy signals.

The noisy signals discussed in this section may be classified in two distinct ways. On the one hand, we shall speak of spatial versus temporal noise. Examples of temporal noise, or random processes where the independent variable is time, include the video signal in a fluoroscopy system, the voltage output from an ultrasonic transducer, and the ratemeter output in a nuclear scanner. Spatial noise refers to point-to-point statistical variations in an image. On the other hand, we may classify noisy signals, either temporal or spatial, as being stationary or nonstationary as defined previously.

The distinction between spatial and temporal noise, while obviously important physically, is rather trivial mathematically. Once a result is derived in the temporal case, it may easily be transcribed to the spatial case by replacing the time t with position $\mathbf{r}$ and temporal frequency ν with spatial

* Some exceptions to this statement have been reported in the literature on nonlinear optics, but they appear to be of no concern for radiographic problems. See Sibilia and Bertolotti (1981) and references given there.

frequency ρ. For notational simplicity we shall deal with temporal noise.

The distinction between stationary and nonstationary noise is more profound mathematically. In particular, Fourier theory is the easiest way to treat stationary noise but is not directly applicable in the nonstationary case. Therefore we devote separate sections to stationary and nonstationary noise.

3.5.1 Stationary Noise

Consider a linear shift-invariant filter with temporal impulse response $p(t)$, and let the input to the filter be the stationary random process $\mathbf{w}_{\text{in}}(t)$. The autocorrelation function of $\mathbf{w}_{\text{in}}(t)$, defined by

$$R_{\text{in}}(\tau) = \langle \mathbf{w}_{\text{in}}(t)\mathbf{w}_{\text{in}}(t+\tau)\rangle, \tag{3.192}$$

is presumed to be known. Therefore the input power spectral density $S_{\text{in}}(\nu)$, which is the Fourier transform of $R_{\text{in}}(\tau)$, is also known.

The random process $\mathbf{w}_{\text{in}}(t)$ is converted by the filter into another random process $\mathbf{w}_{\text{out}}(t)$ measured at the output of the filter. That is, when the input to the filter is one sample function $w_{\text{in}}(t)$ of the input process, the output is the *specific* sample function $w_{\text{out}}(t)$ given by

$$w_{\text{out}}(t) = p(t) * w_{\text{in}}(t) = \int_{-\infty}^{\infty} dt'\, p(t')w_{\text{in}}(t-t'). \tag{3.193}$$

The autocorrelation function of the output signal is

$$\begin{aligned} R_{\text{out}}(\tau) &= \langle \mathbf{w}_{\text{out}}(t)\mathbf{w}_{\text{out}}(t+\tau)\rangle \\ &= \left\langle \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt''\, p(t')p(t'')\mathbf{w}_{\text{in}}(t-t')\mathbf{w}_{\text{in}}(t+\tau-t'')\right\rangle. \end{aligned} \tag{3.194}$$

To proceed further, we should like to take the expectation-value brackets inside the double integral. This step can be justified by writing the expectation value in terms of the joint probability density function for the two random variables $\mathbf{w}_{\text{in}}(t-t')$ and $\mathbf{w}_{\text{in}}(t+\tau-t'')$. To simplify the notation, let us temporarily call these two variables $\mathbf{w}'$ and $\mathbf{w}''$, respectively. Equation (3.194) then becomes

$$R_{\text{out}}(\tau) = \int_{\infty} dw' \int_{\infty} dw'' \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt''\, p(t')p(t'')\,\text{pr}[w', w'']w'w'', \tag{3.195}$$

where the integrals over w' and w'' span the entire domain of those variables. We may now reverse the order of integration and write

$$\begin{aligned} R_{\text{out}}(\tau) &= \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt''\, p(t')p(t'') \int_{\infty} dw' \int_{\infty} dw''\, \text{pr}[w', w'']w'w'' \\ &= \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt''\, p(t')p(t'')\langle \mathbf{w}'\mathbf{w}''\rangle. \end{aligned} \tag{3.196}$$

This step of reversing the order of a deterministic integration and an expectation value will often be useful. Basically, such reversal is valid because statistical averaging is itself just a definite integration with fixed limits.

Equation (3.194) is now

$$R_{\text{out}}(\tau) = \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt''\, p(t')p(t'')R_{\text{in}}(\tau - t'' + t'). \tag{3.197}$$

Note that the absolute time t no longer appears; the output of the filter is stationary if the input is, which is a consequence of the assumption that the filter is shift-invariant.

The change of variables $t''' = t'' - t'$ now allows us to rewrite (3.197)

$$R_{\text{out}}(\tau) = \int_{-\infty}^{\infty} dt''' \left[\int_{-\infty}^{\infty} dt'\, p(t')p(t' + t''') \right] R_{\text{in}}(\tau - t'''). \tag{3.198}$$

The integral over t' is recognized as the *deterministic* autocorrelation $p(t''') \star p(t''')$, not to be confused with a statistical autocorrelation (see Sections 3.1.3 and B.4). The integral over t''' is just a convolution, and we have

$$R_{\text{out}}(\tau) = p(\tau) \star p(\tau) * R_{\text{in}}(\tau). \tag{3.199}$$

Although (3.199) is a straightforward prescription for finding $R_{\text{out}}(\tau)$ when $p(t)$ and $R_{\text{in}}(\tau)$ are known, it is often easier to work in the frequency domain. Equations (B.52) and (B.57) allow us to write

$$S_{\text{out}}(\nu) = |P(\nu)|^2 S_{\text{in}}(\nu), \tag{3.200}$$

where $S_{\text{out}}(\nu)$ and $S_{\text{in}}(\nu)$ are the one-dimensional Fourier transforms of R_{out} and R_{in}, respectively.

This important result should be compared to its deterministic counterpart, (2.26), which says that the Fourier transform of the output of a linear system is equal to the transform of the input times the system transfer function P. If we had an expression for a sample function of the input random process, we could use (2.26) to find the corresponding output sample function, but only rarely do we know the form of a specific sample function. More often we must deal with gross statistical averages like $S_{\text{in}}(\nu)$ or $R_{\text{in}}(\tau)$, in which case (3.200) must be applied since we do not have enough information to use (2.26). Note especially that (2.26) involves $P(\nu)$ linearly, while (3.200) involves its squared magnitude.

Although (3.200) gives us $S_{\text{out}}(\nu)$ and an inverse Fourier transform then gives $R_{\text{out}}(\tau)$, we can often be content with less information than this about the output random process. Often a knowledge of the mean and variance

of the output is sufficient. An expression for the mean is easy to find since

$$\langle \mathbf{w}_{\text{out}}(t) \rangle = \langle p(t) * \mathbf{w}_{\text{in}}(t) \rangle = p(t) * \langle \mathbf{w}_{\text{in}}(t) \rangle, \tag{3.201}$$

where the interchange of convolution and statistical averaging is valid by the same kind of argument that led to (3.196). Equation (3.201) can be simplified further since $\langle \mathbf{w}_{\text{in}}(t) \rangle$ is independent of time by our assumption of stationarity. Therefore we have

$$\begin{aligned} p(t) * \langle \mathbf{w}_{\text{in}}(t) \rangle &= \langle \mathbf{w}_{\text{in}}(t) \rangle \int_{-\infty}^{\infty} p(t')\, dt' \\ &= \langle \mathbf{w}_{\text{in}}(t) \rangle P(0). \end{aligned} \tag{3.202}$$

The variance of $\mathbf{w}_{\text{out}}(t)$, denoted σ_{out}^2, is given by

$$\begin{aligned} \sigma_{\text{out}}^2 &= \langle [\mathbf{w}_{\text{out}}(t)]^2 \rangle - \langle \mathbf{w}_{\text{out}}(t) \rangle^2 \\ &= R_{\text{out}}(\tau = 0) - \langle \mathbf{w}_{\text{out}}(t) \rangle^2, \end{aligned} \tag{3.203}$$

where we have used the fact that the second moment of a stationary random process is just the autocorrelation function evaluated at zero shift [see (3.45)]. In terms of the power spectral density,

$$R_{\text{out}}(0) = \int_{-\infty}^{\infty} S_{\text{out}}(\nu)\, d\nu. \tag{3.204}$$

Combining (3.200)–(3.204) then gives

$$\sigma_{\text{out}}^2 = \int_{-\infty}^{\infty} |P(\nu)|^2 S_{\text{in}}(\nu)\, d\nu - [P(0)\langle \mathbf{w}_{\text{in}}(t) \rangle]^2. \tag{3.205}$$

Note that a knowledge of the mean and variance of the input is *not* sufficient to find the mean and variance of the output; the complete autocorrelation function or power spectral density of the input is required in general.

An important special case of (3.205) arises when $S_{\text{in}}(\nu)$ may be regarded as a constant over the range of frequencies for which $|P(\nu)|^2$ is appreciable. In that case the noise is called *white noise*, meaning that it contains equal amounts of all frequencies of interest, and $S_{\text{in}}(\nu)$ may be removed from the integral in (3.205). We then have

$$\sigma_{\text{out}}^2 = S_{\text{in}}(\bar{\nu}) \int_{-\infty}^{\infty} |P(\nu)|^2\, d\nu - [P(0)\langle \mathbf{w}_{\text{in}}(t) \rangle]^2, \tag{3.206}$$

where $\bar{\nu}$ is some frequency within the passband of the filter.

An extension of this argument leads to a simple physical interpretation of the power spectral density, which so far has been defined rather abstractly as the Fourier transform of the autocorrelation function. Consider a filter with a very narrow passband of width $\Delta\nu$ centered at $\nu = \nu_0$, and let its transfer function be

$$P(\nu) = \text{rect}[(\nu - \nu_0)/\Delta\nu]. \tag{3.207}$$

Since $\text{rect}(x) = [\text{rect}(x)]^2$ except for the two isolated points $x = \pm\frac{1}{2}$ (a set of measure zero for the mathematically inclined), we may write

$$\langle \mathbf{w}_{\text{out}}^2 \rangle = \sigma_{\text{out}}^2 + \langle \mathbf{w}_{\text{out}} \rangle^2 = S_{\text{in}}(\nu_0)\,\Delta\nu. \tag{3.208}$$

The quantity $S_{\text{in}}(\nu_0)$ is thus $(\Delta\nu)^{-1}$ times the second moment of the output signal from this narrowband filter. The reason for the name "power spectral density" can now be explained also. If $\mathbf{w}_{\text{in}}(t)$ is a voltage across a 1-Ω resistor, $\langle \mathbf{w}_{\text{in}}^2 \rangle$ is the average power dissipated in the resistor and S_{in} is the density of this power along the frequency axis, or average power per unit frequency.

3.5.2 Nonstationary Noise

The treatment of nonstationary noise begins in much the same way as the treatment of stationary noise. We are given a linear, shift-invariant system with impulse response $p(t)$. The input is the random process $\mathbf{w}_{\text{in}}(t)$ and the output is the random process $\mathbf{w}_{\text{out}}(t)$. The nonstationary autocorrelation function for the output is

$$R_{\text{out}}(t, t+\tau) = \langle \mathbf{w}_{\text{out}}(t)\mathbf{w}_{\text{out}}(t+\tau) \rangle \times \left\langle \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt''\, p(t')p(t'')\mathbf{w}_{\text{in}}(t-t')\mathbf{w}_{\text{in}}(t+\tau-t'') \right\rangle \tag{3.209}$$

Interchanging the order of integration and statistical averaging and writing the result in terms of the input autocorrelation function produces

$$R_{\text{out}}(t, t+\tau) = \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt''\, p(t')p(t'')R_{\text{in}}(t-t', t+\tau-t''). \tag{3.210a}$$

This equation can be rewritten in shorthand

$$R_{\text{out}}(t, t+\tau) = p(t) * p(t+\tau) * R_{\text{in}}(t, t+\tau). \tag{3.210b}$$

This expression might seem to flout the rules introduced in Appendix B for the convolution shorthand. There we noted that an expression like $f(t_1) * g(t_2)$ would have no meaning if $t_1 \neq t_2$, since we would not know what to take for the output variable or "shift" in the convolution integral. However, the expression $p(t_1) * p(t_2) * R_{\text{in}}(t_1, t_2)$ does not pose this problem since $R_{\text{in}}(t_1, t_2)$ is a function of two variables t_1 and t_2. In the convolution with $p(t_2)$, t_1 is held constant and t_2 is the shift. Conversely, in the convolution with $p(t_1)$, t_2 is held constant. That we write $t_1 = t$ and $t_2 = t + \tau$ is of no consequence; t_1 and t_2 are still independent variables. Any confusion that might be engendered by the shorthand form of (3.210b) can always be resolved by returning to the integral form of (3.210a).

It is also worth noting that (3.210) reduces to its stationary counterpart (3.197) if the input autocorrelation function is indeed stationary, for in that case $R_{\text{in}}(t-t', t+\tau-t'')$ in the integrand of (3.210) may be written

$$R_{\text{in}}[(t+\tau-t'') - (t-t')] = R_{\text{in}}(\tau - t'' + t'). \tag{3.211}$$

Since there are two independent time variables in (3.210), the only way we could obtain a frequency-domain representation would be to use a double Fourier transform with two independent frequency variables. Similarly, a frequency-domain treatment of nonstationary two-dimensional spatial noise would require a fourfold Fourier transform and a four-dimensional frequency space. Fortunately, this degree of complexity is not necessary in radiographic imaging. All the important results can be obtained from (3.210) directly without ever visiting the frequency domain.

The fundamental simplification that occurs in radiographic imaging is that the input noise can almost always be treated as *uncorrelated* (or, more precisely, delta correlated). This term is the nonstationary counterpart of the term *white noise* introduced in Section 3.5.1. If noise is white, the power spectral density is nearly constant and its Fourier transform, the stationary autocorrelation function, approximates a delta function. In the nonstationary case, we cannot speak of a power spectral density, but the autocorrelation function may still be sharply peaked with respect to the τ variable and therefore approximate $\delta(\tau)$.

An important example of uncorrelated noise is the sequence of Poisson impulses introduced in Section 3.3. We showed there that the autocorrelation function for the random process $\mathbf{z}(t)$ defined in (3.116) is

$$R_z(t, t+\tau) = a(t)\,\delta(\tau) + a(t)a(t+\tau), \tag{3.212}$$

where [see (3.142) and (3.145)]

$$\langle \mathbf{z}(t) \rangle = a(t). \tag{3.213}$$

Now let us assume that $\mathbf{z}(t)$ is the input to our linear system. The output correlation function is, from (3.210) and (3.212),

$$\begin{aligned} R_{\text{out}}(t, t+\tau) &= \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt''\, p(t')p(t'')a(t-t')\,\delta(\tau - t'' + t') \\ &\quad + \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt''\, p(t')p(t'')a(t-t')a(t+\tau-t'') \\ &= \int_{-\infty}^{\infty} dt'\, p(t')p(t'+\tau)a(t-t') + [p(t) * a(t)][p(t+\tau) * a(t+\tau)]. \end{aligned} \tag{3.214}$$

The first term is the convolution of the function $p(t)p(t+\tau)$ with $a(t)$. One might be tempted to write $[p(t)p(t+\tau)] * a(t)$, but this form is subject

to misinterpretation. A better shorthand would be $f(t) * a(t)$, where $f(t) = p(t)p(t+\tau)$.

The time-dependent variance of the output signal is given by

$$\sigma^2_{\text{out}}(t) = \langle[\mathbf{w}_{\text{out}}(t)]^2\rangle - \langle\mathbf{w}_{\text{out}}(t)\rangle^2 = R_{\text{out}}(t,t) - \langle\mathbf{w}_{\text{out}}(t)\rangle^2. \tag{3.215}$$

However,

$$\langle\mathbf{w}_{\text{out}}(t)\rangle = p(t) * a(t). \tag{3.216}$$

Therefore

$$\sigma^2_{\text{out}}(t) = \int_{-\infty}^{\infty} dt' \,[p(t')]^2 a(t-t') = [p(t)]^2 * a(t). \tag{3.217}$$

This is by no means a general result. We have used the uncorrelated nature and the Poisson statistics of the input $\mathbf{z}(t)$ in arriving at (3.217). Nevertheless, the result will prove broadly applicable when we treat noise in radiographic images in Chapter 10.

As an example of the formalism developed in this section, let us return to the idealized count-rate meter shown in Fig. 3.1. The detector and count-rate meter may be regarded as a filter whose input is the Poisson random process $\mathbf{z}(t)$ with a nonstationary rate $a(t)$. The impulse response of the filter is assumed to be

$$p(t) = \beta \,\text{rect}[(t - t_d)/T], \tag{3.218}$$

where t_d is a fixed time delay necessary to insure causal behavior, T the averaging time ($T = 1$ sec in Fig. 3.1), and β a constant with dimensions of voltage.

To find the autocorrelation function of the output random process $\mathbf{V}(t)$, note that

$$p(t)p(t+\tau) = \begin{cases} \beta^2 & \text{if } (t_d - \tfrac{1}{2}T) < t < (t_d - \tau + \tfrac{1}{2}T) \\ 0 & \text{otherwise,} \end{cases} \tag{3.219}$$

provided also that τ is in the range $0 < \tau < T$ (see Fig. 3.6).

The autocorrelation function of $\mathbf{V}(t)$ is then, by (3.214),

$$R_V(t, t+\tau) = \beta^2 \int_{t_d - T/2}^{t_d - \tau + T/2} a(t-t')\,dt' + \beta^2 \left(\int_{t_d - T/2}^{t_d + T/2} a(t-t')\,dt' \right) \left(\int_{t_d - T/2}^{t_d + T/2} a(t - t'' + \tau)\,dt'' \right). \tag{3.220}$$

The last term in (3.220) is just $\langle\mathbf{V}(t)\rangle\langle\mathbf{V}(t+\tau)\rangle$. The interesting part of (3.220) is the first term. If, as in Section 3.3.4, we define a zero-mean random process

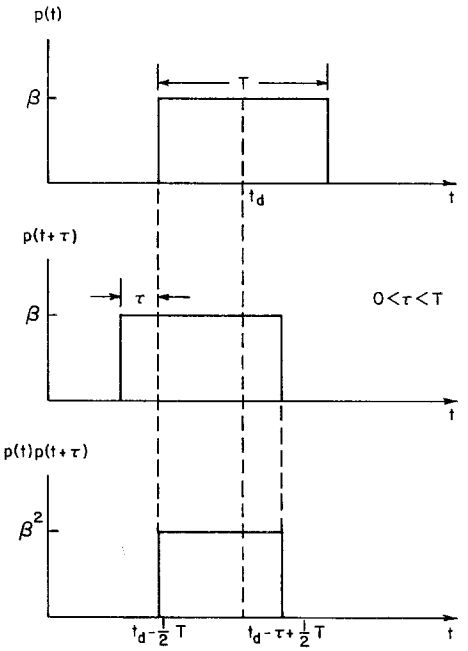

Fig. 3.6 Diagram to aid in the interpretation of (3.219).

$\Delta\mathbf{V}(t)$ by

$$\Delta\mathbf{V}(t) = \mathbf{V}(t) - \langle\mathbf{V}(t)\rangle, \tag{3.221}$$

then we have [cf. (3.147)]

$$R_{\Delta V}(t, t+\tau) = \beta^2 \int_{t_d - T/2}^{t_d - \tau + T/2} a(t-t')\,dt' \qquad (0 < \tau < T). \tag{3.222}$$

This equation shows that, whatever the functional form of $a(t)$, the random process $\Delta\mathbf{V}(t)$ becomes uncorrelated when τ exceeds T since then the range of integration is zero. The filter has a "memory" for a time T, but fluctuations at two times separated by more than T are independent.

The variance of $\mathbf{V}(t)$ is easily calculated from (3.217):

$$\sigma^2_V(t) = [p(t)]^2 * a(t) = \beta^2 \int_{t_d - T/2}^{t_d + T/2} a(t-t')\,dt' = \beta\langle\mathbf{V}(t)\rangle, \tag{3.223}$$

where the last step follows from (3.216) and (3.218). Note that, because of the constant β, the variance does not equal the mean. However, even if β were

unity, $\mathbf{V}(t)$ would still not be a Poisson random process; only nonnegative integer random variables can be Poisson, and $\mathbf{V}(t)$ is a voltage.

Nevertheless, Poisson statistics are an important part of this problem. The time-dependent signal-to-noise ratio is given by

$$\begin{aligned}\mathrm{SNR}(t) &= \langle \mathbf{V}(t)\rangle/\sigma_V(t) = \beta^{-1/2}\langle \mathbf{V}(t)\rangle^{1/2} \\ &= \beta^{-1/2}\left[\beta \int_{t_d-T/2}^{t_d+T/2} a(t-t')\,dt'\right]^{1/2} \\ &= \left[\int_{t-t_d-T/2}^{t-t_d+T/2} a(t'')\,dt''\right]^{1/2}, \end{aligned} \tag{3.224}$$

where the last step follows from the substitution $t'' = t - t'$, with a reversal of the limits of integration to cancel the minus sign in $dt'' = -dt'$. The constant β no longer appears in this result, and the integral has a simple interpretation. It is the total number of counts collected over a time interval of duration T centered about $t = t_d$, where t is the time at which the SNR is measured. The parameter t_d is just the delay inherent in the filter; we do not get an instantaneous measure of the count rate, but must wait a time t_d. However, as expected, the signal-to-noise ratio is simply the square root of the mean number of counts detected in the averaging interval T. This number is a Poisson random variable even if $\mathbf{V}(t)$ is not.

3.5.3 Optimum Filters

From the results of Sections 3.5.1 and 3.5.2, we now know how to calculate the statistical properties of the output of a filter when the statistics of the input and the PSF or MTF of the filter are known. We have not yet addressed the problem of how to choose the "best" filter response. In general, there are four components to this problem (Davenport and Root, 1958): (1) the purpose for which the system is intended; (2) the nature of the input to the filter; (3) the definition of "best"; and (4) the freedom to be allowed the filter designer. A full treatment of this subject is beyond the scope of this book, but we shall briefly consider three specific optimal filters—the Wiener filter, the matched filter, and the Wiener–Helstrom filter.

Wiener Filter

The Wiener filter is intended for the extraction of a signal from additive, signal-independent noise. That is, the input to the filter is the random process

$$\mathbf{y}(t) = \mathbf{s}(t) + \mathbf{n}(t), \tag{3.225}$$

where $\mathbf{s}(t)$ is the signal of interest and $\mathbf{n}(t)$ is noise. In the Wiener formulation (Wiener, 1933, 1949), both $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are regarded as random processes.

We are to form an estimate $\hat{\mathbf{s}}(t)$ of the signal on the basis of the data $\mathbf{y}(t)$. Somewhat arbitrarily, we shall decide that $\hat{\mathbf{s}}(t)$ is to be obtained by passing $\mathbf{y}(t)$ through a linear shift-invariant filter and that the "best" estimate will be the one that minimizes the mean-squared error:

$$\hat{\mathbf{s}}(t) = \mathbf{y}(t) * p(t), \tag{3.226}$$

$$\langle |\hat{\mathbf{s}}(t) - \mathbf{s}(t)|^2 \rangle = \text{minimum}, \tag{3.227}$$

where $p(t)$ is the impulse response of the filter. We further assume that the signal and noise are uncorrelated,

$$\langle \mathbf{s}(t)\mathbf{n}(t')\rangle = 0 \qquad \text{for all} \quad t \text{ and } t', \tag{3.228}$$

and that we know the autocorrelation functions for both $\mathbf{s}(t)$ and $\mathbf{n}(t)$.

One of the most elegant solutions to this problem is based on the *orthogonality principle*. Papoulis (1965) has shown that the $p(t)$ that minimizes $\langle |\hat{\mathbf{s}}(t) - \mathbf{s}(t)|^2 \rangle$ is also the one that makes the error $\hat{\mathbf{s}}(t) - \mathbf{s}(t)$ "orthogonal" to the data, i.e.,

$$\langle [\hat{\mathbf{s}}(t) - \mathbf{s}(t)]\mathbf{y}(t')\rangle = 0 \qquad \text{for all} \quad t \text{ and } t'. \tag{3.229}$$

Use of (3.226) and (3.228) in (3.229) yields

$$\begin{aligned}&\langle \{[\mathbf{s}(t) * p(t)] + [\mathbf{n}(t) * p(t)] - \mathbf{s}(t)\}\{\mathbf{s}(t') + \mathbf{n}(t')\}\rangle \\ &\quad = \int_{-\infty}^{\infty} \langle \mathbf{s}(t'')\mathbf{s}(t')\rangle p(t-t'')\,dt'' \\ &\qquad + \int_{-\infty}^{\infty} \langle \mathbf{n}(t'')\mathbf{n}(t')\rangle p(t-t'')\,dt'' - \langle \mathbf{s}(t)\mathbf{s}(t')\rangle = 0. \end{aligned} \tag{3.230}$$

Each of these expectation values is just an autocorrelation function, and we have

$$\int_{-\infty}^{\infty} R_s(t''-t')p(t-t'')\,dt'' + \int_{-\infty}^{\infty} R_n(t''-t')p(t-t'')\,dt'' = R_s(t-t'), \tag{3.231}$$

where we have assumed that both $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are stationary. The change of variables $\tau = t'' - t'$ shows that

$$\begin{aligned}\int_{-\infty}^{\infty} R_s(t''-t')p(t-t'')\,dt'' &= \int_{-\infty}^{\infty} R_s(\tau)p(t-t'-\tau)\,d\tau \\ &= R_s(t-t') * p(t-t'), \end{aligned} \tag{3.232}$$

and similarly for the integral involving $R_n(t''-t')$. Therefore (3.231) can be Fourier-transformed to yield

$$S_s(\nu)P(\nu) + S_n(\nu)P(\nu) = S_s(\nu), \tag{3.233}$$

where S_s and S_n are the power spectral densities of $\mathbf{s}$ and $\mathbf{n}$. The optimum filter thus has a transfer function given by

$$P(\nu) = \frac{S_s(\nu)}{S_s(\nu) + S_n(\nu)}. \tag{3.234}$$

This result is intuitively appealing since it shows that $P(\nu)$ should be large for frequencies at which the signal power is much larger than the noise power, and that $P(\nu)$ should be small if the noise power is the larger.

Matched Filter

While the Wiener filter is optimal for a certain class of *estimation* problems, the matched filter is used in *detection* problems. In these problems, the interest is not in accurately reconstructing an estimate of an unknown signal, but rather in detecting the presence of a known signal (Klauder *et al.*, 1960; Turin, 1960).

The input to the filter is *either* signal plus noise or signal alone:

$$\mathbf{y}(t) = s(t) + \mathbf{n}(t) \qquad \text{or} \qquad \mathbf{y}(t) = \mathbf{n}(t), \tag{3.235}$$

where now $s(t)$ is a known function rather than a random variable.

The output of the filter is $\mathbf{y}^\dagger(t)$, given by

$$\mathbf{y}^\dagger(t) = \mathbf{y}(t) * p(t). \tag{3.236}$$

We use the notation $\mathbf{y}^\dagger(t)$ rather than $\hat{\mathbf{s}}(t)$ since the output may not bear any resemblance at all to the input. Instead, we want the filter to give us a peak or recognition spike if the signal is present. Since we have the freedom to specify the overall time delay of the filter, we can choose to have this peak occur at an arbitrary time t_1. A reasonable criterion to optimize the filter is then to maximize the signal-to-noise ratio of the output at $t = t_1$, i.e.,

$$[\mathrm{SNR}(t_1)]^2 = \frac{[s(t_1) * p(t_1)]^2}{\langle [\mathbf{n}(t_1) * p(t_1)]^2 \rangle} = \text{maximum}, \tag{3.237}$$

where we are assuming that $\mathbf{n}(t)$ has zero mean. The denominator of (3.237) is just the noise variance at the filter output. By (3.205) it is given (for stationary noise) by

$$\begin{aligned} \langle [\mathbf{n}(t_1) * p(t_1)]^2 \rangle &= \sigma_{\text{out}}^2 \\ &= R_{\text{out}}(0) = \int_{-\infty}^{\infty} S_n(\nu) |P(\nu)|^2 \, d\nu. \end{aligned} \tag{3.238}$$

If the noise is white, $S_n(\nu)$ is constant over the bandwidth of the filter, and

$$\sigma_{\text{out}}^2 = S_n \int_{-\infty}^{\infty} |P(\nu)|^2 \, d\nu. \tag{3.239}$$

The numerator of (3.237) is

$$[s(t_1) * p(t_1)]^2 = \left| \int_{-\infty}^{\infty} \exp(2\pi i \nu t_1) S(\nu) P(\nu) \, d\nu \right|^2, \tag{3.240}$$

where $S(\nu)$, the Fourier transform of $s(t)$, should not be confused with a power spectral density.

It is convenient to normalize the signal part of the SNR by the total signal power $\int_{-\infty}^{\infty} |S(\nu)|^2 \, d\nu$, and to normalize the noise part by S_n. Neither of these normalizations has any effect on the choice of $p(t)$ since both normalizing factors are independent of $p(t)$; a filter that maximizes the normalized SNR will also maximize the unnormalized one. We thus require that

$$\frac{S_n |\mathrm{SNR}(t_1)|^2}{\int_{-\infty}^{\infty} |S(\nu)|^2 \, d\nu} = \text{maximum}, \tag{3.241}$$

which, with (3.239) and (3.240), becomes

$$\frac{\left| \int_{-\infty}^{\infty} \exp(2\pi i \nu t_1) S(\nu) P(\nu) \, d\nu \right|^2}{\left[\int_{-\infty}^{\infty} |S(\nu)|^2 \, d\nu \right] \left[\int_{-\infty}^{\infty} |P(\nu)|^2 \, d\nu \right]} = \text{maximum}. \tag{3.242}$$

We now appeal to the well-known Schwarz inequality (Klauder *et al.*, 1960), which states that two arbitrary complex functions (of a real variable) $u(\nu)$ and $v(\nu)$ must satisfy

$$\frac{\left| \int_{-\infty}^{\infty} u(\nu) v(\nu) \, d\nu \right|^2}{\left[\int_{-\infty}^{\infty} |u(\nu)|^2 \, d\nu \right] \left[\int_{-\infty}^{\infty} |v(\nu)|^2 \, d\nu \right]} \leqslant 1. \tag{3.243}$$

The equality holds only if $u(\nu)$ is proportional to $[v(\nu)]^*$, which in the present problem means that

$$P(\nu) \propto [S(\nu)]^* \exp(-2\pi i \nu t_1). \tag{3.244}$$

After an inverse transform, this condition is equivalent to

$$p(t) \propto s(t_1 - t). \tag{3.245}$$

In other words, $p(t)$ is a shifted and time-reversed replica of the signal we are trying to detect.

The time reversal has the effect of converting the convolution integral $\mathbf{y}(t) * p(t)$ into a correlation integral as defined in (B.53). If t_1 is zero, which

is possible only in noncausal systems, we have

$$\begin{aligned}\mathbf{y}(t) * p(t) &= \int_{-\infty}^{\infty} \mathbf{y}(t')p(t-t')\,dt' \\ &= \int_{-\infty}^{\infty} \mathbf{y}(t')s(t'-t) = \mathbf{y}(t) \star s(t).\end{aligned} \tag{3.246}$$

Matched filtering is thus equivalent to correlating the data with the signal we are trying to detect.

Applications of matched filtering are discussed in Chapters 8 and 10.

Wiener-Helstrom Filter

An important variation on the Wiener filter was devised by Helstrom (1967). He considered the problem of estimating a signal that has been corrupted in two ways—by being convolved with a known filter function (usually a blurring or low-pass filter) and by the addition of signal-independent noise. The data to be filtered are of the form

$$\mathbf{y}(t) = [\mathbf{s}(t) * h(t)] + \mathbf{n}(t), \tag{3.247}$$

where $h(t)$ and the autocorrelation functions of $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are presumed known. An estimate $\hat{\mathbf{s}}(t)$ of the signal is to be formed by filtering $\mathbf{y}(t)$ with a filter of impulse response $p(t)$. Again we choose the minimum mean-squared error as the optimality condition.

Helstrom showed that the optimum filter for this problem has a transfer function given by

$$P(v) = \frac{H^*(v)}{|H(v)|^2 + [S_n(v)/S_s(v)]}. \tag{3.248}$$

Several limits are of interest. First, note that if $h(t) = \delta(t)$, $H(v) = 1$, and the original Wiener filter is recovered. Second, if the signal-to-noise ratio is very good at all frequencies, such that $[S_n(v)/S_s(v)]^2 \ll |H(v)|^2$, then $P(v)$ reduces to an inverse filter, $P(v) = [H(v)]^{-1}$. Finally, if the SNR is very poor and both signal and noise are white, then $S_n(v)/S_s(v)$ is large and constant and the Wiener–Helstrom filter becomes a matched filter $P(v) \propto [H(v)]^*$.