Fractal-Based Point Processes

2005

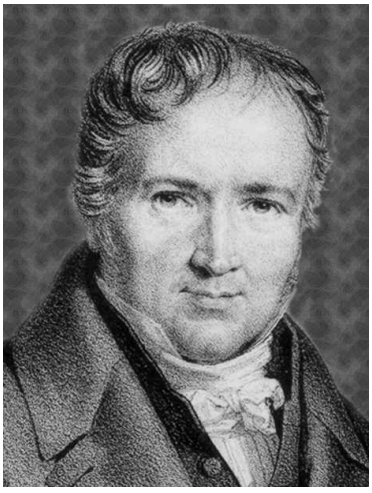
Steven Bradley Lowen
Harvard Medical School
McLean Hospital

Malvin Carl Teich
Boston University
Columbia University

WILEY

3

Point Processes: Definition and Measures



The famous French mathematician **Siméon Denis Poisson (1781–1840)** developed the fundamental probability distribution that bears his name; its applications are legion in a broad variety of fields.

Together with Major Greenwood, the Scottish engineer **George Udny Yule (1871–1951)** conceived an important generalization of the Poisson distribution in which the rate itself becomes a random variable.

3.1 POINT PROCESSES

Some random phenomena occur at discrete times or locations, with the individual events largely identical. Examples include the events of a radioactive decay process (Sec. 2.5.4), vehicles passing a certain location on a road (Fig. 1.2), the arrival of information packets at a node of a computer communication network (Chapter 13), the occurrence of action potentials in a neural preparation (Chapter 5), and the occurrences of QRS complexes in the electrocardiogram (Sec. 12.2.2).

In all of these cases, the set of times at which the events occur comprises the salient characteristics of the process. The details of the events themselves are less important, inasmuch as one event closely resembles another. A **stochastic point process**, often abbreviated as **point process**, is a mathematical construct that represents these events as random points in a space. We use the terms "event" and "point" interchangeably.

Point-process theory grew out of studies in a number of fields: population processes, cosmic-ray showers, component durability, and queueing problems in communications engineering. The theory of point processes took shape as a discipline in the 1920s, and the literature in this area grew rapidly in the following decades (Lubberger, 1925, 1927; Lotka, 1939; Campbell, 1939; Fréchet, 1940; Palm, 1943; Feller, 1948; Wold, 1948, 1949; Bartlett, 1955; Moyal, 1962). Writing in German, the Swedish mathematician Palm (1943) coined the term *Punktprozesse*: "point process."

A comprehensive bibliography detailing some of the early landmarks of point-process theory and analysis, in the general context of stochastic processes, is available (Wold, 1965). A concise early history of the field appears in Daley & Vere-Jones (1988, Chapter 1).

Many modern books on the topic adopt a rigorous and abstract approach (Brillinger, 1981; Leadbetter, Lindgren & Rootzen, 1983; Daley & Vere-Jones, 1988; Kingman, 1993; Reiss, 1993; Baccelli & Brémaud, 2003), which has the merit of providing a great deal of generality. Other books on point processes are more didactic and applications oriented (Parzen, 1962; Cox, 1962; Cox & Lewis, 1966; Feller, 1971; Lewis, 1972; Srinivasan, 1974; Saleh, 1978; Cox & Isham, 1980; Snyder & Miller, 1991), offering specific examples useful in the physical and biological sciences. We adopt a rather informal approach to point processes, and concentrate particularly on those that exhibit fractal characteristics.

Some point processes depend on space as well as time; lightning strikes, for example, deliver more electrical activity to some areas than to others. However, we confine the treatment provided here to one-dimensional point processes; other dimensions, if present, are not incorporated into the model.

Since a variety of time variables exist, we adopt the following conventions: (1) lowercase roman italic letters (generally t) refer to absolute time, measured with respect to an origin that does not depend on the point process under study; (2) uppercase roman italic letters (generally T) refer to a duration over which events are analyzed — again, the duration does not depend on the point process under study and the analysis need not begin at the origin. Finally, the symbol τ generally represents the times between events.

3.2 REPRESENTATIONS

Figure 3.1 presents several representations useful in the analysis of point processes (Teich, Heneghan, Lowen & Turcott, 1996). Panel a) demonstrates a realization of a point process as a series of impulses occurring at specified times t_n. Since these impulses have vanishing width, they are most rigorously defined in terms of the derivative of a well-defined counting process $N(t)$ [panel b)], a monotonically increasing function of t, which starts at the origin and augments by unity when an event occurs. Accordingly, we write the point process itself as $dN(t)$, to emphasize its strict definition within the context of an integral. The point-process representation thus belongs to the family of generalized functions (Bracewell, 1986).

The set of **event times** $\{t_n\}$, or equivalently the **sequence of interevent intervals** $\{\tau_n\}$ (together with t_0), where $\tau_n = t_{n+1} - t_n$, completely describe the point process.[1] Furthermore, the **sequence of counts** depicted in Fig. 3.1c) also contains much information about the process. Here we divide the time axis into equally spaced contiguous counting durations of T sec to produce a sequence of counts $\{Z_k(T)\}$,

[1] To remove ambiguity, we define $N(t)$ as a right-continuous process, so that $N(t_n) = n$.

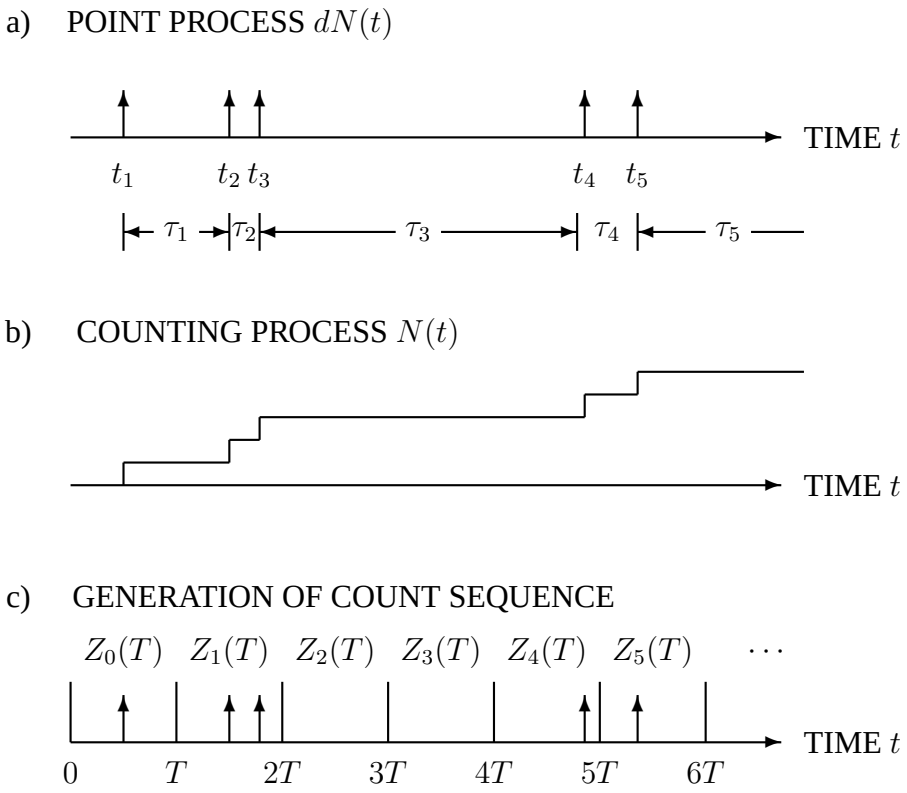


Fig. 3.1 Representations of a point process. (a) A sequence of idealized impulses, occurring at times t_n, represents the events, and form a stochastic point process $dN(t)$. We also show the interevent intervals $\tau_n = t_{n+1} - t_n$. For convenience of analysis, several alternative representations of the point process appear. (b) The counting process $N(t)$ begins at a value of zero at $t = 0$. At every event occurrence the value of $N(t)$ augments by unity. (c) The sequence of counts $\{Z_k(T)\}$, a discrete-time nonnegative integer-valued stochastic process, derives from the point process by recording the number of events in successive counting durations of length T.

where $Z_k(T) = N[(k+1)T] - N(kT)$ denotes the number of events in the kth duration. As illustrated in panel d), this sequence forms a discrete-time random process of nonnegative integers. In general, forming the sequence of counts loses information, although for an orderly point process (see below) decreasing the size of the counting duration T reduces the loss to an arbitrarily small value. An attractive feature of this representation lies in the fact that it preserves the correspondence between the discrete time axis of the counting process $\{Z_k(T)\}$ and the absolute "real" time axis of the underlying point process. Within the process of counts $\{Z_k(T)\}$, the elements $Z_k(T)$ and $Z_{k+n}(T)$ refer to the number of counts in durations separated by precisely

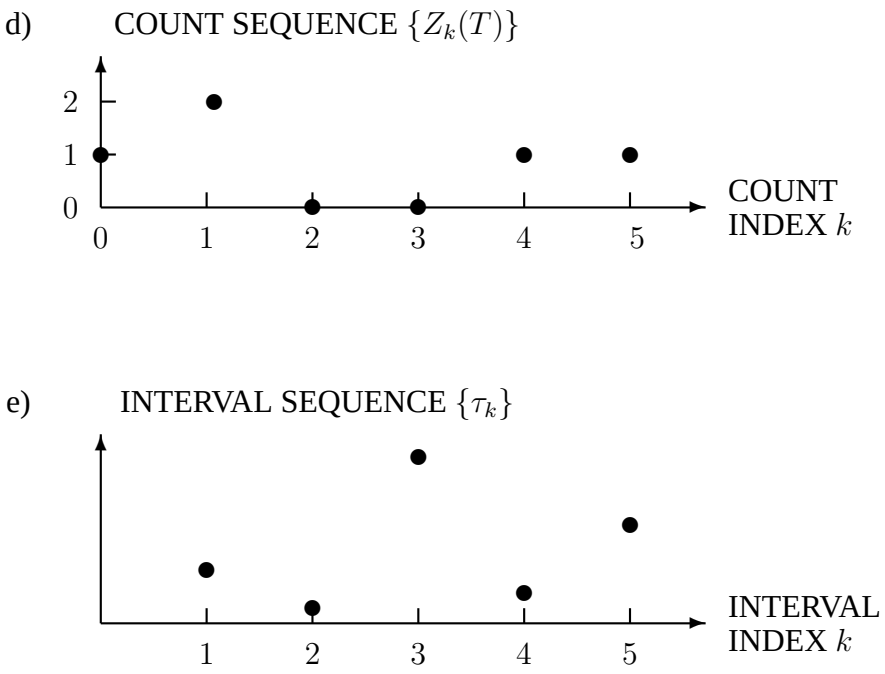


Fig. 3.1 (continued) (d) The sequence of counts $\{Z_k(T)\}$ depends on a count index k. The counting process destroys information because this representation eliminates the precise timing of events within each counting duration. Correlations in the discrete-time sequence $\{Z_k(T)\}$ can be readily interpreted in terms of real time. (e) The sequence of interevent intervals $\{\tau_k\}$ represents the time between successive events, yielding a discrete-time, positive, real-valued stochastic process. All information contained in the original point process remains in this representation, but the discrete-time axis of the sequence of interevent intervals suffers random distortion relative to the real time axis of the point process.

$(n-1)T$ sec, so that we can readily associate correlation in the process $\{Z_k(T)\}$ with correlation in the underlying point process $dN(t)$.

Much as the sequence of counts forms an auxiliary process, so does the sequence of interevent intervals. Figure 3.1e) presents the intervals $\{\tau_k\}$ drawn from the point process in panel a), indexed by interval number. In contrast to the sequence of counts, this representation preserves all of the information of the point process $dN(t)$, but eliminates the direct correspondence between absolute time and the index number. The sequence of intervals therefore affords only rough comparisons with correlations in the underlying point process, particularly for intervals with a large coefficient of variation [see Eq. (3.5)].

We restrict ourselves largely to **orderly point processes**, which essentially means that no two events occur at the same time, and that events do not localize to any single

time. Formally speaking, in terms of the counting process $N(t)$ we have

$$\lim_{\epsilon \to 0} \epsilon^{-1} \Pr\{N(t+\epsilon) - N(t) > 1\} = 0 \tag{3.1}$$

for any time t, which also implies the lack of coincident events (Daley, 1974). We also generally consider stationary processes; all statistics remain unchanged despite any shifting of the time axis,

$$\begin{aligned} &\Pr\{f[N(s_1), N(s_2), \ldots, N(s_k)] < x\} = \\ &\quad \Pr\{f[N(s_1+s), N(s_2+s), \ldots, N(s_k+s)] < x\}, \end{aligned} \tag{3.2}$$

for any arbitrary real-valued function $f(\cdots)$ with any number k of arguments and any offset time s. Unless explicitly stated otherwise, all point processes in this book are orderly and stationary.

For some applications, point processes as defined above do not suffice for all sets of events; customers arriving at a queue, for example, may require widely different service times (see Chapter 13). While still strongly localized, the events do not then resemble each other. In addition to its location, each event then requires a descriptive mark, such as the service time in this example. A generalization of the point-process model to a **marked point process** version accommodates problems of this type (Matthes, 1963; Cox & Isham, 1980; Sigman, 1995). All marks in a marked point process are of the same type (examples of types include integers, real numbers, vectors, and functions). Including marks adds additional information (but also complexity) to point-process models. Since most of the point processes considered in this book do not warrant this level of effort, we largely restrict ourselves to unmarked point processes.

We now proceed to describe a number of measures that prove useful in the study of point processes. In general, no one statistic, or even small group of statistics, suffices to completely characterize a point process; each provides a different view of the process and highlights different properties. A good description requires many such views. The statistics fall into three broad classes: those based on the intervals between events, as displayed in Fig. 3.1e); those based on the counting process, as shown in Fig. 3.1d); and those based on the point process as a whole, as depicted in Fig. 3.1a).

3.3 INTERVAL-BASED MEASURES

Conversion of a point process $dN(t)$ into a sequence of intervals between events $\{\tau_k\}$ reduces $dN(t)$ to a discrete-time real-valued process, for which a wide variety of analytical methods exist.

3.3.1 Marginal statistics

Perhaps the simplest statistics of a point process ignore any dependencies among event times, and focus on the marginal properties of the interevent intervals $\{\tau_k\}$. These fall into two classes.

The first includes the **probability distribution**, the **survivor function**, and the **probability density** where this derivative exists:

$$\begin{array}{lrcl} \text{distribution} & P_\tau(t) & = & \Pr\{\tau \le t\} \\ \text{survivor} & \mathsf{S}_\tau(t) & = & \Pr\{\tau > t\} \;=\; 1 - P_\tau(t) \\ \text{density} & p_\tau(t) & = & dP_\tau(t)/dt. \end{array} \tag{3.3}$$

For a well-defined point process, we require that $P_\tau(t) = p_\tau(t) = 0$ for $t < 0$.

The second class comprises the **moments** $\mathrm{E}[\tau^n]$ and the statistics derived therefrom, such as:

$$\begin{array}{ll} \text{variance} & \mathrm{Var}[\tau] = \mathrm{E}[\tau^2] - \mathrm{E}^2[\tau] \\ \text{standard deviation} & \sigma_\tau = \sqrt{\mathrm{Var}[\tau]} \\ \text{skewness} & \mathrm{E}[(\tau - \mathrm{E}[\tau])^3]/\sigma_\tau^3 \\ \text{kurtosis} & \mathrm{E}[(\tau - \mathrm{E}[\tau])^4]/\sigma_\tau^4 - 3, \end{array} \tag{3.4}$$

where these moments exist.[2] The interval **coefficient of variation** C_τ, a commonly used measure of the relative dispersion (relative width) of the intervals, is defined as

$$C_\tau \equiv \sigma_\tau / \mathrm{E}[\tau]. \tag{3.5}$$

As with all random variables, the **characteristic function** $\phi_\tau(\omega)$, defined as[3]

$$\phi_\tau(\omega) \equiv \int_0^\infty p_\tau(t)\, e^{-i\omega t}\, dt, \tag{3.6}$$

forms a compact representation of the moments. In general we have

$$i^n \frac{d^n}{d\omega^n}\, \phi_\tau(\omega)_{\omega=0} = \mathrm{E}[\tau^n] \tag{3.7}$$

[2] Several definitions for skewness exist (such as the difference between the mean and the median, all divided by the standard deviation), and some authors define kurtosis without the "3" subtracted. All definitions in general use have their merits, and no clear winner emerges. However, we choose to make use of the particular forms provided above for three reasons. First, the skewness and kurtosis provided in Eq. (3.4) are given by normalized versions of the third and fourth cumulants or semi-invariants, respectively. Second, these versions prove most analytically tractable. And finally, both of these definitions assume a value of zero for a Gaussian random variable.

[3] Some define the characteristic function with the argument of the exponential $i\omega t$ rather than $-i\omega t$. Since both i and $-i$ form equally valid square roots of -1, it is convention, rather than mathematics, that determines the choice. We employ the expression shown in Eq. (3.6) because it leads to simpler results.

for the moments, and

$$i^n \frac{d^n}{d\omega^n} \ln\big[\phi_\tau(\omega)\big]_{\omega=0} = C_n \tag{3.8}$$

for the **cumulants** or **semi-invariants** C_n. The moments and cumulants determine each other, and in particular we have

$$\begin{aligned}
\mathrm{E}[\tau] &= C_1 \\
\mathrm{E}[\tau^2] &= C_2 + C_1^2 \\
\mathrm{E}[\tau^3] &= C_3 + 3C_1C_2 + C_1^3 \\
\mathrm{E}[\tau^4] &= C_4 + 4C_1C_3 + 3C_2^2 + 6C_1^2C_2 + C_1^4 \\
\mathrm{Var}[\tau] &= C_2 \\
\text{skewness} &= C_3/C_2^{3/2} \\
\text{kurtosis} &= C_4/C_2^2.
\end{aligned} \tag{3.9}$$

In addition to the interevent-interval statistics, the **forward recurrence time** also proves useful. This represents the time $\vartheta(t)$ remaining to the next event of a point process, starting at an arbitrary time t independent of the process. Formally, we have

$$\vartheta(t) = t_k - t, \quad \text{where} \quad k = N(t) + 1. \tag{3.10}$$

A simple relation exists between probability distribution functions of the interevent time and the forward recurrence time (see Prob. 3.8)

$$P_\vartheta(s) = \Pr\{\vartheta(t) \le s\} = \frac{1}{\mathrm{E}[\tau]} \int_0^s [1 - P_\tau(x)]\, dx, \tag{3.11}$$

which yields the statistics of $\vartheta(t)$ through Eqs. (3.3) and (3.4). In particular, taking the derivative of Eq. (3.11) yields

$$p_\vartheta(s) = [1 - P_\tau(s)]/\mathrm{E}[\tau]. \tag{3.12}$$

Thus, a normalized version of the interevent-interval survivor function provides the recurrence-time probability density.

While the interevent-interval probability distribution and survivor function exist for all point processes, some moments may not; in particular, for a probability density function that decays as $t^{-\alpha}$ for large t, moments $\mathrm{E}[\tau^n]$ for $n \ge \alpha - 1$ will not exist. For example, the density function

$$p_\tau(t) = \sqrt{t_0/\pi}\ t^{-3/2} \exp(-t_0/t), \qquad t > 0, \tag{3.13}$$

with t_0 a fixed positive parameter, has infinite moments $\mathrm{E}[\tau^n]$ for all positive integers n (Feller, 1971) (see Prob. 3.6).

Densities such as these belong to the family of **heavy-tailed distributions**, for which

$$\lim_{t\to\infty} \frac{\mathsf{S}_\tau(t + t_1)}{\mathsf{S}_\tau(t)} = 1, \qquad t_1 \ge 0, \tag{3.14}$$

for any fixed, finite time t, where again $\mathsf{S}_\tau(t) = 1 - P_\tau(t)$ is the interval survivor function. **Subexponential distributions**, introduced by Chistyakov (1964), form an important subclass of heavy-tailed distributions (see, for example, Embrechts, Klüppelberg & Mikosch, 1997; Sigman, 1999; Greiner, Jobmann & Klüppelberg, 1999). These distributions have survivor functions $\mathsf{S}_\tau(t)$ that obey

$$\lim_{t\to\infty} e^{\epsilon t}\, \mathsf{S}_\tau(t) = \infty, \qquad \epsilon > 0, \tag{3.15}$$

so that the tail of the survivor function tends to zero more slowly than any exponential function $e^{-\epsilon t}$. Examples of subexponential distributions include the Pareto and its variants, the lognormal, and the stretched exponential (Weibull[4]).

For a particular class of point processes, called **renewal point processes** (see Sec. 4.2), the values of each interevent interval do not depend on those before or after it. For this class of point processes only, the marginal statistics described above, and in particular the probability distribution function alone, determine the entire behavior of the processes. Generally, however, dependencies do occur among interevent intervals and this necessitates the use of several statistics for an overview of the sequence of intervals, and of the point process itself.

3.3.2 Interval autocorrelation

The **interval autocorrelation** $R_\tau(k)$, which provides further information about point processes that do not belong to the renewal point-process family, is defined as

$$R_\tau(k) \equiv \mathrm{E}[\tau_n \tau_{n+k}]\,. \tag{3.16}$$

For independent intervals, $R_\tau(k) = \mathrm{E}^2[\tau]$ for $k \neq 0$, confirming that the autocorrelation then provides no additional information.

A normalized version of this measure proves useful in many cases. Subtracting the value returned for independent intervals, $\mathrm{E}^2[\tau]$, and dividing by the interval variance, $\mathrm{Var}[\tau]$, yields the **interval serial correlation coefficient**

$$\varrho_\tau(k) \equiv \frac{R_\tau(k) - \mathrm{E}^2[\tau]}{\mathrm{Var}[\tau]}\,. \tag{3.17}$$

By construction, $\varrho_\tau(0) = 1$ for any point process. For independent intervals, $\varrho_\tau(k) = 0$ for $k \neq 0$.

Inasmuch as no direct relationship generally exists between the lag variable k and time t in seconds, these measures, as well as the other interval-based measures that follow, have limited usefulness. This restriction is relaxed when the mean interval greatly exceeds the interval standard deviation (DeBoer, Karemaker & Strackee, 1984), in which case $t \approx k\,\mathrm{E}[\tau]$ (see the beginning of Sec. 3.4).

[4]The Weibull distribution follows the form $P_\tau(t) = 1 - \exp[-(t/t_0)^\xi]$, typically with a shape parameter $0 < \xi < 1$ (see Gumbel, 1958, pp. 279, 302); the exponential distribution is recovered for $\xi = 1$. This distribution possesses finite moments of all orders but is nevertheless heavy-tailed because the survivor function decays more slowly than any exponential.

3.3.3 Interval spectrum

Fourier transforming the autocorrelation in Eq. (3.16) yields the **interval-based spectrum** $S_\tau(\mathsf{f})$:

$$S_\tau(\mathsf{f}) = \sum_k R_\tau(k)\, e^{-i2\pi k \mathsf{f}}, \tag{3.18}$$

where f is the (dimensionless) frequency with units of cycles per number of intervals. For independent intervals, $S_\tau(\mathsf{f}) = \mathrm{Var}[\tau]$ for all $\mathsf{f} \neq 0$. Again, the independent variable f has no simple connection with its conventional counterpart (the frequency f in Hz), so this measure principally finds use in processes with small deviations from periodicity. The performance of a normalized version of this statistic is examined in Sec. 12.3.7.

3.3.4 Interval wavelet variance

A particularly appropriate method for characterizing signals with fractal behavior is via the use of wavelets (Daubechies, 1992). Fourier analysis, employed in generating the interval spectrum, decomposes a signal into a series of basis functions, all of which have different shapes: sinewaves of varying frequency, phase, and amplitude but identical duration. The basis functions differ in the number of cycles they contain. Wavelet decomposition, in contrast, employs basis functions that all have the same shape, and derive from a prototype wavelet by stretching and shifting. As with their Fourier counterpart, the inversion of wavelet transforms to return the original signal proves relatively simple.

This self-affine basis set makes wavelets well suited for analyzing signals that contains statistical copies of themselves; signals that exhibit fractal characteristics in time fall into this category. Wavelet-based methods for characterizing fractal signals yield estimates superior to those obtained by many other methods (Thurner et al., 1997; Abry, Flandrin, Taqqu & Veitch, 2000, 2003, see also Chapter 12), and wavelet analysis also enjoys the salutary property of removing nonstationarities from the signal under study (Teich et al., 1996; Abry & Flandrin, 1996; Arneodo, Grasseau & Holschneider, 1988).

The **wavelet transform** of a sequence of interevent intervals takes the form (Daubechies, 1992; Aldroubi & Unser, 1996; Akay, 1997; Abry et al., 2003)

$$W_{\psi,\tau}(k,l) = \sum_n 2^{-k/2}\, \psi(2^{-k}n - l)\, \tau_n, \tag{3.19}$$

where the continuous-time wavelet function $\psi(x)$ satisfies a number of admissibility criteria (Daubechies, 1992). We consider the **interval wavelet-transform variance** in Eq. (3.19), since the transform itself is a random variable. As a result of one of the admissibility criteria, wavelet transforms have zero mean, so that

$$\begin{aligned} &\mathrm{Var}[W_{\psi,\tau}(k,l)] \\ &\quad = \quad \mathrm{E}[W^2_{\psi,\tau}(k,l)] \end{aligned}$$

$$= \mathrm{E}\left[\sum_n \sum_m 2^{-k}\,\psi(2^{-k}n-l)\,\psi(2^{-k}m-l)\,\tau_m\,\tau_n\right]$$
$$= 2^{-k}\sum_n \sum_m \psi(2^{-k}n-l)\,\psi(2^{-k}m-l)\,R_\tau(m-n). \qquad (3.20)$$

The performance of a normalized version of this measure, defined as

$$A_\tau(k) \equiv \frac{\mathrm{Var}[W_{\psi,\tau}(k,l)]}{\mathrm{Var}[\tau]}\,, \qquad (3.21)$$

is examined in Sec. 12.3.6.

For stationary point processes, the wavelet variance does not depend on the position index l, but only on the scale variable k. Under these conditions, the interval wavelet variance is directly related to the interval spectrum (see Sec. 3.3.3) via an integral transform (Heneghan, Lowen & Teich, 1999). Knowledge of one of these measures is thus equivalent to knowledge of the other. The wavelet variance exhibits the same nonlinear relationship between the lag variable k and conventional time as observed above for several other measures.

3.3.5 Rescaled range analysis

Harold Hurst studied the water flow patterns of the river Nile, and discovered the presence of long-term fluctuations in the yearly flood levels.[5] He observed that years with greater-than-average flow tended to cluster together, as did years with lower-than-average flow, but that no characteristic cluster size appeared to exist. Hurst (1951) developed **rescaled range analysis** (R/S analysis) to quantify this effect, and this statistic became the first robust method for characterizing fractal behavior in discrete-time sequences.[6] The rescaled range statistic provides information about dependencies among interevent intervals (or other sequences) in a form fundamentally different from that obtained by the interval autocorrelation $R_\tau(k)$.

One algorithm for calculating the rescaled range proceeds as follows. Begin by selecting a set of k interevent intervals that start with the first available interval. From this set, estimate the mean $\widehat{\mathrm{E}}[\tau]$ and (biased) standard deviation

$$\sqrt{\frac{k-1}{k}\,\widehat{\mathrm{Var}}[\tau]}\,. \qquad (3.22)$$

Next, subtract the estimated mean, divide by the biased standard deviation, and construct a running sum of this rescaled process. Now generate the rescaled range by

[5] Hurst (1956) and Hurst, Black & Simaika (1965) studied the water flow through other rivers as well; they also examined variations in other natural time series such as rainfall, temperature, pressure, tree-ring thickness, and sunspot numbers.

[6] The unexpected clustering results, together with the seemingly *ad hoc* character of the rescaled range statistic, led to a lack of acceptance of Hurst's work that lasted until he was in his seventies (Mandelbrot, 1982, pp. 396–398). A photograph of Hurst appears at the beginning of Chapter 12 and a biographical sketch is provided by Mandelbrot (1982, Chapter 40).

subtracting the minimum value that the sum attains from its maximum. Next, repeat this procedure for all possible contiguous blocks of k values within the entire data set, and average these values together to yield the rescaled-range estimate $\widehat{U}(k)$. Finally, repeat this procedure for a variety of lags k. Figure 3.2 provides pseudocode for this algorithm, and Fig.3.3 presents a schematic graphical calculation.

For independent intervals, we have

$$U(k) \approx \sqrt{k}, \tag{3.23}$$

where the exact relationship depends on the distribution of the $\{\tau_n\}$ (see Prob. 3.7).

While this measure now enjoys broad popularity for the study of processes that exhibit long-term correlation or large moments (Hurst, 1951; Feller, 1951; Hurst, 1956; Hurst et al., 1965; Mandelbrot, 1982; Mandelbrot & Wallis, 1969c,b; Schepers, van Beek & Bassingthwaighte, 1992), it suffers from large systematic errors for some sequences (Beran, 1994; Bassingthwaighte & Raymond, 1994; Caccia, Percival, Cannon, Raymond & Bassingthwaighte, 1997). Nevertheless, it can prove useful in some cases since it robustly handles data sets with infinite variance (Mandelbrot, 2001, Chapter 5, pp. 155–171). In Sec. 12.3.4 we examine the performance of a normalized form of this the rescaled range statistic,

$$U_2(k) \equiv U^2(k)/k\,. \tag{3.24}$$

Calculate R/S from discrete-time sequence $\{x_n\}$ of length M:
  set $k = 2$
  while M/k large /* typically require $M/k \geq 10$ */
    set $m = 0$
    while $m + k - 1 \leq M$
      estimate mean $\widehat{\mathrm{E}}[x] = \sum_{n=m}^{m+k-1} x_n/k$
      estimate biased std. dev. $\widehat{\sigma}_x^2 = \sum_{n=m}^{m+k-1} \left(x_n - \widehat{\mathrm{E}}[x]\right)^2 / k$
      generate normalized sequence: $y_n = \left(x_n - \widehat{\mathrm{E}}[x]\right) / \widehat{\sigma}_x$
      generate summed sequence: $z_n = z_{n-1} + y_n; \quad z_1 = y_1$
      find minimum and maximum values:
        $z_{\min} = \min(z_n), \quad m \leq n < m + k$
        $z_{\max} = \max(z_n), \quad m \leq n < m + k$
      find the difference between them: $U(k, m) = z_{\max} - z_{\min}$
      increment starting index: $m \to m + 1$
                  /* $m \to m + k$ faster, almost as accurate */
    end while
    average all values of $U(k, m)$ to yield $\widehat{U}(k)$
    report k and $\widehat{U}(k)$
    increase block size k /* typically $k \to 2k$ */
  end while

Fig. 3.2 Rescaled-range analysis: Pseudocode.

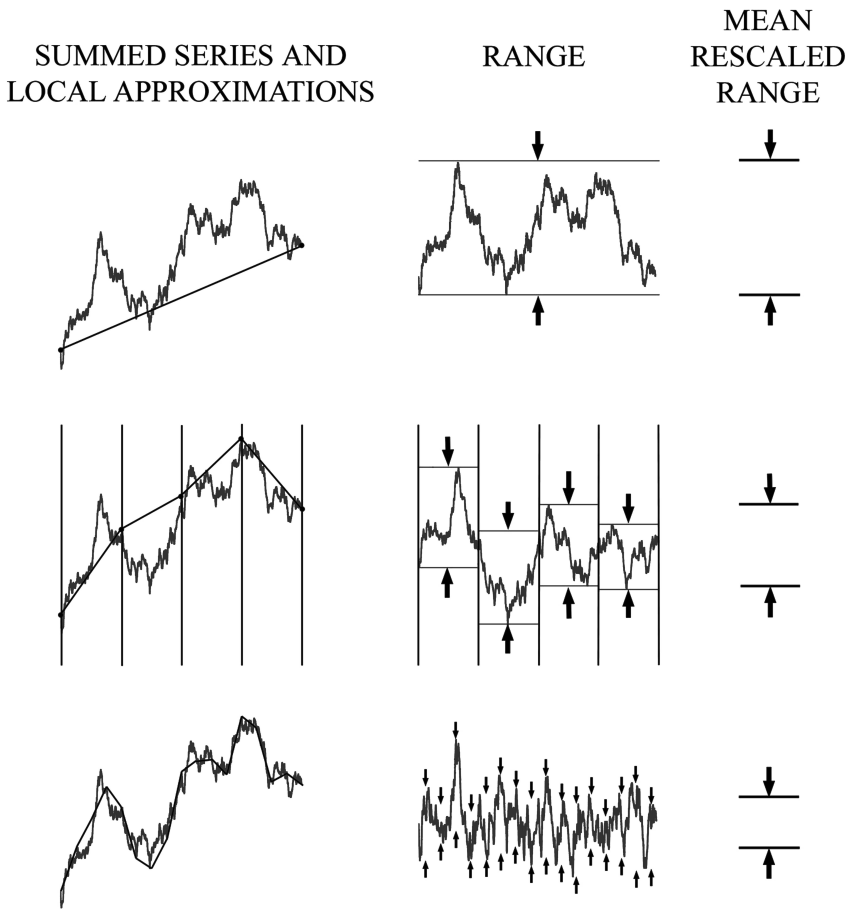


Fig. 3.3 Top row: a summed series and its approximation by a linear function through its first and last values (left column). The difference between the summed series and the linear function yields a range (middle column). After rescaling by the (biased) sample standard deviation, this single range value forms the mean rescaled range (right column). Middle row: the process repeats with the same summed series divided into four subseries, each with its own linear function, as well as range and standard deviation. Their average forms the mean rescaled range. Bottom row: the same calculation for sixteen subseries.

3.3.6 Detrended fluctuation analysis

Detrended fluctuation analysis offers yet another method for analyzing dependencies among interevent intervals (Peng et al., 1995). The algorithm for calculating the detrended fluctuation begins by constructing a running sum of the interval sequence over all indices. Next divide the summed series into blocks of length k and perform a least-squares fit on each of the data blocks, providing the trends for the individual

blocks. Now detrend the sequence by subtracting the local trend in each block. Next, sum the squares of the detrended fluctuations, divide by k, and take the square root to obtain the detrended fluctuation estimate $\widehat{Y}(k)$. Finally, repeat this procedure for a variety of lags k. Figure 3.4 provides pseudocode for this algorithm, and Fig.3.5 presents a sample graphical calculation.

Like $U(k)$, for independent intervals $Y(k)$ varies as $\sqrt{k}$ for large k (see Prob. 3.7); more precisely,

$$Y(k) = \sigma_\tau \sqrt{(k^2 - 4)/15\,k} \tag{3.25}$$

for all $k > 2$ (see Sec. A.1.1). Except for the special case of Gaussian-distributed sequences, detrended fluctuation analysis exhibits significant bias and variance (Taqqu & Teverovsky, 1998). Section 12.3.5 reports the performance of a normalized form of the detrended fluctuation statistic,

$$Y_2(k_2) \equiv \frac{15\,Y^2(k+2)}{(k+2)\,\mathrm{Var}[\tau]}\,. \tag{3.26}$$

An extension of detrending exists, which involves the removal of higher-order polynomial trends rather than merely linear ones (Hu, Ivanov, Chen, Carpena & Stanley, 2001). Like the use of wavelets with several vanishing moments, this renders detrended fluctuation analysis insensitive to such trends in the data. Much as the interval wavelet variance is directly related to the interval spectrum, as discussed in Sec. 3.3.4, an analytical link exists between detrended fluctuation analysis and the interval spectrum (Heneghan & McDarby, 2000). Similar relations also connect the normalized count-based variance, the normalized count-based wavelet variance, and the conventional spectrum, as shown later in this chapter.

Calculate detrended fluctuation analysis

```
from discrete-time sequence {x_n} of length M:
  generate summed sequence:  y_n = y_{n-1} + x_n;   y_1 = x_1
  set k = 3                                          /* or set k = 4 */
  while M/k large                          /* typically require M/k ≥ 10 */
    set m = 0
    while m + k - 1 ≤ M
      in this mth block of k values in {y_n},
      find least-squares linear fit z_n = an + b
        to y_n over the range mk < n ≤ (m+1)k
      subtract the fit from the summed sequence:  w_n = y_n - z_n
      sum the squares of the remainder:  q_m = Σ_{n=mk+1}^{(m+1)k} w_n^2
      normalize the sum:  v_m = q_m/k
      take the square root:  Y(k,m) = √v_m
      increment block index:  m → m + 1
    end while
    average all values of Y(k,m) to yield Ŷ(k)
    report k and Ŷ(k)
    increase block size k                        /* typically k → 2k */
  end while
```

Fig. 3.4 Detrended fluctuation analysis: Pseudocode.

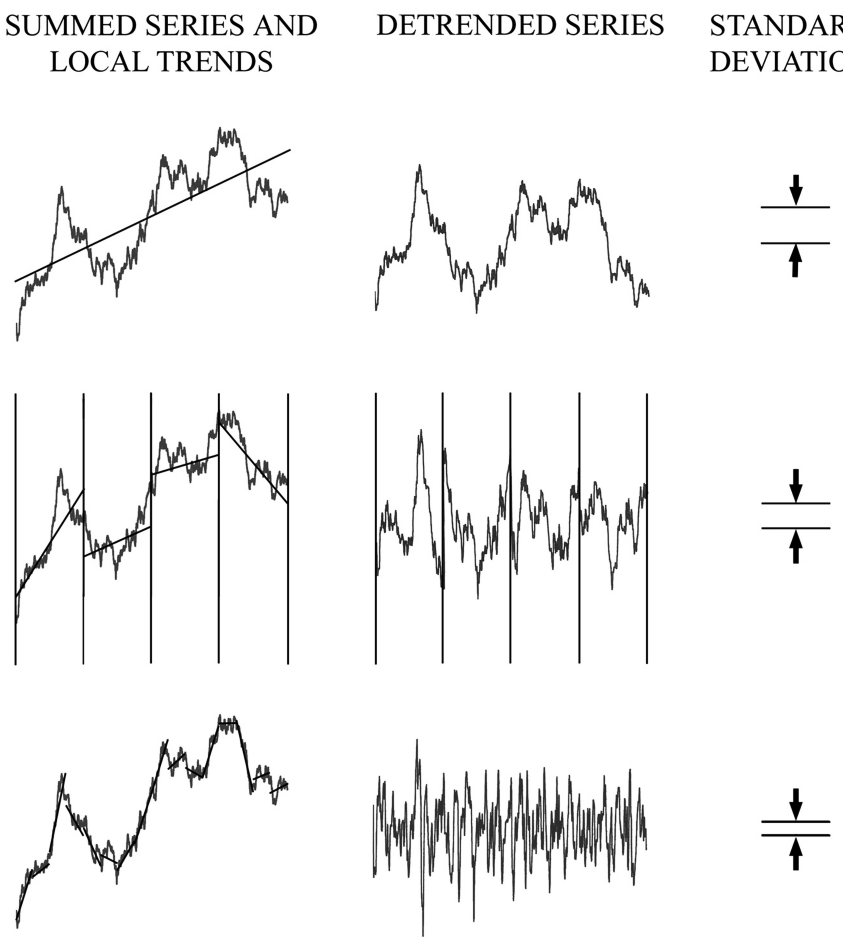


Fig. 3.5 Top row: a summed series and its least-squares fit by a linear function (left column). The difference between the summed series and the linear function (middle column). The sample standard deviation of this detrended series (right column). Middle row: the process repeats with the same summed series divided into four subseries, each with its own linear function as well as detrended series and sample standard deviation. Bottom row: the same calculation for sixteen subseries.

3.4 COUNT-BASED MEASURES

Count-based measures form the second broad class of point-process statistics, and derive from the sequence of counts $\{Z_k(T)\}$ shown in Fig. 3.1d). Perhaps the earliest application of a stochastic *counting* process was offered in the domain of legal decisions, by Poisson himself in 1837. Other early applications were Seidel's (1876)

analysis of thunderstorms, blood-cell counting via microscopy carried out by Abbe (1878), Rutherford & Geiger's (1910) famous α-particle counting experiments (see also Bateman, 1910), and the well-known study of accident proneness conducted by Greenwood & Yule (1920). Researchers sometimes favor count-based measures because the information they provide corresponds to the real time of the point process. Moreover, counting measures can be used for systems that intrinsically involve integration. Count-number statistics also provide the only systematic analysis approach available for spaces of dimension greater than one.

Rate-based measures form a closely related set, where the **sample rate** $\lambda_k(T)$ is a normalized version of the sequence of counts:

$$\lambda_k(T) = Z_k(T)/T. \tag{3.27}$$

A variant of the counting procedure, called **generalized rates**, assigns fractional counts, depending on the manner in which an interevent interval spans different counting durations (Papoulis, 1991). Referring to Fig. 3.1c), for example, an interevent interval begins just before time $t = 2T$ and extends to just before $t = 5T$. Applying the generalized rate method for this case, one would set $\lambda_2(T) = \lambda_3(T) \approx 0.3/T$. Similarly, the duration from $t = T$ to $t = 2T$, with sample rate denoted $\lambda_1(T)$, would contain a small portion of the long interval discussed above (perhaps a tenth of it), plus about half of the interval that spans the divider $t = T$, plus the entire interval that spans the two events located within the duration from $t = T$ to $t = 2T$. Thus, one would set $\lambda_1(T) \approx (0.1 + 0.5 + 1)/T = 1.6/T$ using this method.

The generalized rates yield somewhat smoother estimates of the rate by reducing the quantization noise inherent in constructing the sequence of counts; they find their principal use in point processes for which the mean interval greatly exceeds the interval standard deviation, such as in heartbeat sequences (DeBoer et al., 1984). A relationship between count- and interval-based measures can be formulated in this special case, as indicated in connection with the second-order interval-based measures in Secs. 3.3.2 and 3.3.3. Since the events follow a relatively regular spacing for these point processes, results from one domain (count- or interval-based) can be readily translated into those in the other domain via the relation $t \approx k\,\mathrm{E}[\tau]$ (DeBoer et al., 1984).

While some measures, such as the spectrum, appear extensively in the literature as both interval- and count-based versions, others typically do not, despite being theoretically possible. For example, both rescaled range and detrended fluctuation analyses could serve as count-based measures, but published studies have traditionally not included such analyses in count-based form.

3.4.1 Marginal statistics

As with interval-based measures, we begin with count-based statistics that ignore any dependencies among counts, focusing on the marginal properties of the counts $\{Z_k(T)\}$ instead. These again fall into two classes. The **counting distribution** (or **probability mass function**), which is akin to the probability density function for

discrete-valued random variables, forms the first:

$$p_Z(n;T) = \Pr\{Z(T) = n\}. \tag{3.28}$$

We employ the shorthand notation $p_Z(n)$ where this does not introduce confusion.

If no events occur in a time of duration T, then the time to the next event must exceed T. These two descriptions of the same outcome must have identical probabilities, so that (see Prob. 3.4)

$$\Pr\{Z(t) = 0\} = 1 - P_\vartheta(t) \tag{3.29}$$

where $P_\vartheta(t)$ again denotes the forward-recurrence-time probability distribution function. Combining Eqs. (3.11) and (3.29) yields (see Prob. 3.3)

$$p_\tau(t) = \mathrm{E}[\tau]\,\frac{d^2}{dt^2}\Pr\{Z(t) = 0\}, \tag{3.30}$$

thereby forming a connection between the interval-based and count-based domains.

For processes $N(t)$, and counting times T for which the number of counts greatly exceeds unity, the corresponding rate $\lambda(T)$ can assume any of a large number of possible values. In this case, the individual values of the counting distribution for $Z(T)$ [and therefore for $\lambda(T)$ as well] all become small. A continuous approximation for $\lambda(T)$, and its description by a probability density function, then becomes reasonable.

The second class of marginal count-based measures comprises the **count moments** $\mathrm{E}[Z^n(T)]$, and the statistics derived from them, such as the following:

$$\begin{array}{ll} \text{variance} & \mathrm{Var}[Z] = \mathrm{E}[Z^2] - \mathrm{E}^2[Z] \\ \text{factorial moments} & \mathrm{E}[Z!/(Z-k)!] \\ \text{skewness} & \mathrm{E}[(Z-\mathrm{E}[Z])^3]/\mathrm{Var}^{3/2}[Z] \\ \text{kurtosis} & \mathrm{E}[(Z-\mathrm{E}[Z])^4]/\mathrm{Var}^2[Z] - 3, \end{array} \tag{3.31}$$

where these moments exist, and where we suppress the explicit reference to the counting time T (see Footnote 2 on p. 55). Here $k! \equiv k\cdot(k-1)\cdots 3\cdot 2\cdot 1$ represents the factorial function, and we employ the notational convenience that $1/n! \equiv 0$ for n a negative integer.

For renewal point processes, as discussed in Sec. 4.2, the value of each interevent interval does not depend on those before or after it. However, for general renewal point processes the *count* sequence *does* exhibit dependence. For example, consider a point process constructed so that the smallest possible interval takes a value larger than $\tau_{\min}$. At a counting time half as large ($T = \tau_{\min}/2$), the observation that $Z_{k-1}(T) = 1$ implies that $Z_k(T) = 0$; otherwise the interval spanning $t = kT$ would not exceed $2T = \tau_{\min}$. The only point process for which the sequence of counts $\{Z_k(T)\}$ exhibits independence for all counting times T is the homogeneous Poisson process described in Sec. 4.1. For all other point processes, dependencies do occur among counts, necessitating the use of several statistics for an overview of the sequence of counts.

3.4.2 Normalized variance

As the counting duration T increases, all of the count moments provided in Eq. (3.31) increase, suggesting that some manner of normalization might prove useful. The **normalized variance** $F(T)$ is obtained by dividing the variance by the mean:

$$F(T) \equiv \frac{\mathrm{Var}[Z(T)]}{\mathrm{E}[Z(T)]}. \tag{3.32}$$

This quantity appears to have been first devised by Ugo Fano (1947) to characterize the statistical fluctuations of the number of ions generated by individual fast charged particles; it therefore garnered the appellation "Fano factor." It also goes by a number of other names including "count variance-to-mean ratio," "dispersion ratio," and "index of dispersion" (Cox & Isham, 1980). We prefer the terminology "normalized variance" for its simplicity and apt description.

Figure 3.6a) displays its construction. In general, the normalized variance is a function of the counting time. In the limit of small counting times T, the count $Z(T)$ generally takes a value of zero, and rarely unity; larger values occur with negligible frequency. The result then becomes a sequence of **Bernoulli trials**. Under these conditions,

$$\begin{aligned} \Pr\{Z(T)=1\} &= p_Z(1) \\ \Pr\{Z(T)=0\} &= p_Z(0) \approx 1-p_Z(1) \\ \Pr\{Z(T)>1\} &= \sum_{n>1} p_Z(n) \approx 0, \end{aligned} \tag{3.33}$$

so that (see Prob. 3.2)

$$\begin{aligned} \lim_{T\to 0} F(T) &= \lim_{T\to 0} \frac{\mathrm{Var}[Z(T)]}{\mathrm{E}[Z(T)]} = \lim_{T\to 0} \frac{\mathrm{E}[Z^2(T)] - \mathrm{E}^2[Z(T)]}{\mathrm{E}[Z(T)]} \\ &\approx \lim_{T\to 0} \frac{\mathrm{E}[Z(T)] - \mathrm{E}^2[Z(T)]}{\mathrm{E}[Z(T)]} = \lim_{T\to 0} \big\{1 - \mathrm{E}[Z(T)]\big\} \\ &\approx \lim_{T\to 0} [1 - p_Z(1)] = 1 \end{aligned} \tag{3.34}$$

for any orderly point process. The normalized variance approaches unity, as expected for a sequence of Bernoulli trials. For the homogeneous Poisson process, $F(T) = 1$ for all counting times, as shown in Sec. 4.1.

As an estimator for finite-length data sets, the normalized variance suffers from bias for sequences of counts that exhibit dependence (Lowen & Teich, 1995; Thurner et al., 1997), as we demonstrate in Sec. 12.3.2. In contrast, the normalized Haar-wavelet variance described in the next section does not have this limitation.

3.4.3 Normalized Haar-wavelet variance

In the same way that wavelets can prove useful in characterizing the sequence of intervals derived from a fractal or fractal-rate point process (see Sec. 3.3.4), they also

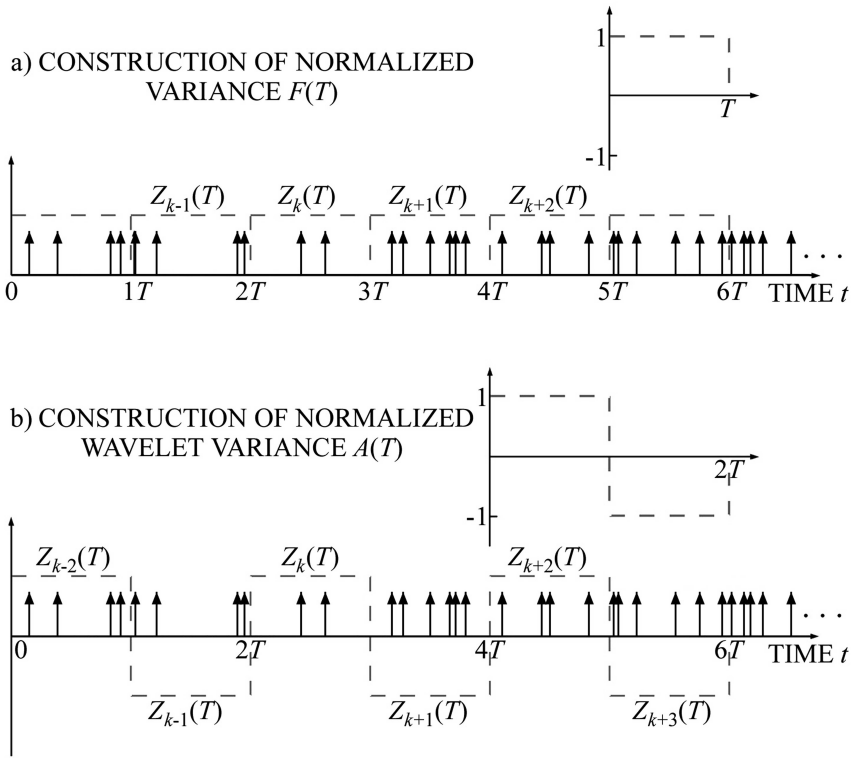


Fig. 3.6 A point process gives rise to a sequence of counts $\{Z_k(T)\}$ by counting the number of events in each contiguous time duration T. (a) Computing the variance of the number of counts, and dividing this quantity by the mean number of counts, yields the normalized variance $F(T)$. (b) Computing the variance of the *difference* in the number of counts in adjacent counting durations, and dividing this quantity by twice the mean number of counts, yields the normalized Haar-wavelet variance $A(T)$.

provide an appropriate method for the analysis of count sequences associated with such processes.

In terms of the point process, the wavelet transform becomes

$$C_{\psi,N}(a,b) = \int a^{-1/2}\,\psi[(t-b)/a]\,dN(t), \tag{3.35}$$

where the continuous wavelet transform now applies. We focus on a particular wavelet, the Haar wavelet (1910), defined by

$$\psi_{\mathrm{Haar}}(t) = \begin{cases} 1 & \text{for } 0 \le t < \frac{1}{2} \\ -1 & \text{for } \frac{1}{2} \le t < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{3.36}$$

Equation (3.35) then becomes

$$\begin{aligned} C_{\mathrm{Haar},N}(a,b) &= a^{-1/2}\int_b^{b+a/2} dN(t) - a^{-1/2}\int_{b+a/2}^{b+a} dN(t) \\ &= a^{-1/2}\Big\{[N(b+a/2) - N(b)] \\ &\qquad - [N(b+a) - N(b+a/2)]\Big\}. \end{aligned} \tag{3.37}$$

Setting $T = a/2$ and $k = 2b/a$ yields

$$C_{\mathrm{Haar},N}(2T,k) = (2T)^{-1/2}\,[Z_k(T) - Z_{k+1}(T)] \tag{3.38}$$

$$\mathrm{E}\big[C^2_{\mathrm{Haar},N}(2T,k)\big] = \frac{\mathrm{E}\big\{[Z_k(T) - Z_{k+1}(T)]^2\big\}}{2T}\,. \tag{3.39}$$

Dividing by the sample rate yields the **normalized Haar-wavelet variance** $A(T)$:

$$\begin{aligned} A(T) &= \frac{\mathrm{E}\big[C^2_{\mathrm{Haar},N}(2T,k)\big]}{\mathrm{E}[\lambda_k(T)]} \\ &= \frac{\mathrm{E}\big\{[Z_k(T) - Z_{k+1}(T)]^2\big\}}{2T}\,\frac{T}{\mathrm{E}[Z_k(T)]} \\ &\equiv \frac{\mathrm{E}\big\{[Z_k(T) - Z_{k+1}(T)]^2\big\}}{2\mathrm{E}[Z_k(T)]}. \end{aligned} \tag{3.40}$$

We schematically illustrate the construction of this quantity from a point process in Fig. 3.6b). We initially developed this measure for the analysis of action-potential sequences recorded from the mammalian auditory nerve (Lowen & Teich, 1996a). At first we called it the "Allan factor" by virtue of its relationship to the Allan variance (Allan, 1966; Barnes & Allan, 1966), but we prefer the appellation "normalized Haar-wavelet variance," which is more descriptive.[7]

Unlike the normalized variance discussed in Sec. 3.4.2, $A(T)$ does not suffer from bias for sequences of counts that exhibit dependence (Abry & Flandrin, 1996; Thurner et al., 1997), as we demonstrate in Secs. 12.2.3, 12.3.8, and 12.4. In terms of the normalized variance, we have (Scharf, Meesmann, Boese, Chialvo & Kniffki, 1995):

$$\begin{aligned} 2F(T) - F(2T) &= 2\frac{\mathrm{Var}[Z(T)]}{\mathrm{E}[Z(T)]} - \frac{\mathrm{Var}[Z(2T)]}{\mathrm{E}[Z(2T)]} \\ &= \frac{4\mathrm{E}[Z^2(T)] - 4\mathrm{E}^2[Z(T)]}{2\mathrm{E}[Z(T)]} \\ &\quad - \frac{\mathrm{E}\{[Z_k(T) + Z_{k+1}(T)]^2\} - \mathrm{E}^2\{[Z_k(T) + Z_{k+1}(T)]\}}{2\mathrm{E}[Z(T)]} \end{aligned}$$

[7] Photographs of Haar and Allan appear at the beginnings of Chapters 5 and 12, respectively.

$$
\begin{aligned}
&= \frac{4\mathrm{E}[Z^2(T)] - 4\mathrm{E}^2[Z(T)]}{2\mathrm{E}[Z(T)]} \\
&\quad + \frac{-2\mathrm{E}[Z^2(T)] - 2\mathrm{E}[Z_k(T)\, Z_{k+1}(T)] + 4\mathrm{E}^2[Z(T)]}{2\mathrm{E}[Z(T)]} \\
&= \frac{\mathrm{E}\{[Z_k(T) - Z_{k+1}(T)]^2\}}{2\mathrm{E}[Z(T)]} \\
&= A(T). \qquad (3.41)
\end{aligned}
$$

In particular,

$$\lim_{T\to 0} A(T) = \lim_{T\to 0} F(T) = 1. \qquad (3.42)$$

Furthermore, if the normalized variance increases monotonically, we have

$$
\begin{aligned}
F(2T) &> F(T) \\
0 &> F(T) - F(2T) \\
F(T) &> F(T) + F(T) - F(2T) \\
F(T) &> A(T), \qquad (3.43)
\end{aligned}
$$

illustrating that the normalized variance always exceeds the normalized Haar-wavelet variance in this case.

3.4.4 Count autocorrelation

In direct parallel to the sequence of intervals, the autocorrelation of the sequence of counts provides information about their second-order properties. The **count autocorrelation** is defined as

$$R_Z(k,T) \equiv \mathrm{E}[Z_n(T) Z_{n+k}(T)]\,. \qquad (3.44)$$

For a sequence of counts that exhibits independence at a counting time T, we have $R_Z(k,T) = \mathrm{E}^2[Z(T)]$ for $k \neq 0$. In contrast to the interval-based autocorrelation in Eq. (3.16), in this case a direct relationship exists between the lag variable k and time t in seconds: $k = t/T$.

As with the count variance, estimates of this measure suffer from bias, as we demonstrate in Sec. 12.3.3 using a normalized form of the autocovariance,

$$R_2(k) \equiv \frac{R_Z(k,T) - \mathrm{E}^2[Z(T)]}{\mathrm{Var}[Z(T)]}\,. \qquad (3.45)$$

A generalized version of the wavelet variance defined by

$$C_{\mathrm{Haar},N}(2T,k) \equiv \mathrm{E}\Big\{[Z_n(T) - Z_{n+k}(T)]^2\Big\} \qquad (3.46)$$

would not suffer from bias, but this quantity has not enjoyed wide use.

3.4.5 Rate spectrum

Again paralleling results for the sequence of intervals, Fourier transforming the autocorrelation in Eq. (3.44) yields the **count-based spectrum** $S_Z(f,T)$. In practice, the **rate-based spectrum**, often simply called the **rate spectrum**, proves more useful:

$$S_\lambda(f,T) = T^{-2} S_Z(f,T) = \frac{1}{T} \sum_k R_Z(k,T)\, e^{-i2\pi k f T}. \tag{3.47}$$

This quantity derives from the Fourier transform of the observed sequence of rates, rather than the sequence of counts, and since $\lambda_k = Z_k(T)/T$, a factor of T^{-2} appears in Eq. (3.47). Because of this normalization, in the range $0 < f \ll 1/T$ the rate spectrum $S_\lambda(f,T)$ approaches a limiting value that does not depend on T. The count-based version, $S_Z(f,T)$, does not enjoy this property, and is therefore less useful.

We examine the performance of $S_\lambda(f,T)$ in Secs. 12.3.9 and 12.4. An estimate of the rate-based spectrum derived from a data set is often referred to as the **periodogram** (this term is also used for an estimate of the interval-based spectrum). One can reduce the variance of the rate-based periodogram by averaging it over nearby frequencies; or by partitioning the original point process into blocks of identical length, and then averaging the individual periodograms thus obtained; or by employing both methods in tandem.

3.5 OTHER MEASURES

Other measures of a point process exist that do not fall into either the interval-based or count-based categories. Some of these measures prove problematical in practice, leading to long computation times, poor statistics, or both; they are, nevertheless, of interest as theoretical constructs.

3.5.1 Coincidence rate

The coincidence rate $G(t)$ measures the correlation between events as a function of a specified time delay t, regardless of any intervening events (Kuznetsov & Stratonovich, 1956; Kuznetsov, Stratonovich & Tikhonov, 1965; Cox & Lewis, 1966):

$$\begin{aligned} G(t) &\equiv \lim_{\epsilon \to 0} \frac{1}{\epsilon^2} \Pr\Big\{ N(s+\epsilon) - N(s) > 0 \\ &\qquad\qquad \text{and } N(s+t+\epsilon) - N(s+t) > 0 \Big\} \qquad (3.48) \\ &= \mathrm{E}\left[\frac{dN(s)}{ds} \, \frac{dN(s+t)}{ds} \right]. \qquad (3.49) \end{aligned}$$

For $t = 0$, the two probabilities coincide in Eq. (3.48), which leads to an infinite value. It proves most convenient to represent this as a Dirac delta function:

$$\lim_{\substack{\epsilon>0\\ \epsilon\to 0}} \int_{-\epsilon}^{\epsilon} G(t)\,dt = \lim_{\substack{\epsilon>0\\ \epsilon\to 0}} \int_{-\epsilon}^{\epsilon} \delta(t)\,\mathrm{E}[\mu]\,dt = \mathrm{E}[\mu]. \tag{3.50}$$

The coincidence rate is the point-process analog of the autocorrelation used for continuous-time processes. For large delays t, the two quantities inside the expectation in Eq. (3.49) become independent, so that

$$\begin{aligned} \lim_{t\to\infty} G(t) &= \lim_{t\to\infty} \mathrm{E}\left[\frac{dN(s)}{ds}\,\frac{dN(s+t)}{ds}\right] \\ &= \mathrm{E}\left[\frac{dN(s)}{ds}\right]\mathrm{E}\left[\frac{dN(s+t)}{ds}\right] \\ &= \mathrm{E}[\mu(s)]\;\mathrm{E}[\mu(s+t)] \\ &= \mathrm{E}^2[\mu], \end{aligned} \tag{3.51}$$

where $\mu(t) \equiv dN(t)/dt$ denotes the **instantaneous rate** of the point process $dN(t)$. In general, $\mu(t)$ varies in a random fashion, representing the local likelihood of event generation. For a stationary point process, however, the corresponding rate has statistics that do not vary with time. In this case (which describes the vast majority of point processes considered in this book), we can eliminate the explicit dependence on the time t for the marginal statistics. For the mean value, for example, we write $\mathrm{E}[\mu]$ instead of $\mathrm{E}[\mu(t)]$. A further simplification takes place in the special case when the history of the process has no effect on the instantaneous generation rate, as for the homogeneous Poisson process described in Sec. 4.1. The expectation itself then becomes superfluous, so we eliminate the explicit expectation operator as well; $\mathrm{E}[\mu]$ then becomes μ. Returning to the general point-process case, we have $\mathrm{E}[\mu] = 1/E[\tau]$, where we interpret $1/\infty$ as zero.

We can conveniently express several of the count-based measures set forth earlier in terms of the coincidence rate $G(t)$ (Cox & Isham, 1980; Thurner et al., 1997). Expressions for the normalized variance $F(T)$ (see Prob. 3.10), normalized Haar-wavelet variance $A(T)$, and autocorrelation $R_Z(k,T)$ are given by

$$F(T) = \frac{1}{\mathrm{E}[\mu]T}\int_{-T}^{T} \left\{G(t) - \mathrm{E}^2[\mu]\right\}(T-|t|)\,dt \tag{3.52}$$

$$A(T) = \frac{2}{\mathrm{E}[\mu]T}\int_{-T}^{T} \left[G(t) - G(2t)\right](T-|t|)\,dt \tag{3.53}$$

$$R_Z(k,T) = \int_{-T}^{T} G(kT+t)\,(T-|t|)\,dt, \tag{3.54}$$

respectively. We can readily invert Eq. (3.52) to yield

$$G(t) = \mathrm{E}[\mu]\,\delta(t) + \mathrm{E}^2[\mu] + \frac{\mathrm{E}[\mu]}{2}\frac{d^2}{dT^2}\left[TF(T)\right]_{T=t}. \tag{3.55}$$

In practice, determining the coincidence rate from a finite set of data proves impossible. The probability that two events exist in the data set with a separation of precisely t is zero, for any *a priori* value of t. Instead, the autocorrelation $R_Z(k,T)$ provides an estimate of the coincidence rate through Eq. (3.54). For T smaller than the time scale over which $G(t)$ varies significantly, we have

$$\begin{aligned} R_Z(k,T) &= \int_{-T}^{T} G(kT+t)\,(T-|t|)\,dt \\ &\approx G(kT)\int_{-T}^{T}(T-|t|)\,dt = G(kT)\,T^2 \\ G(t) &\approx T^{-2}R_Z(t/T,T). \end{aligned} \tag{3.56}$$

However, obtaining useful resolution in this approximation requires a small value of T, which, in turn, leads to excessive variance in this estimator of $G(t)$ for all but the largest data sets. For this reason, the coincidence rate is rarely used in practice.

3.5.2 Point-process spectrum

As with the sequence of intervals $\{\tau_k\}$ and the sequence of counts $\{Z_k(T)\}$, Fourier transformation of the coincidence rate $G(t)$ yields a spectrum $S_N(f)$, this time the spectrum of the point process $dN(t)$ itself[8] (Bartlett, 1963, 1964):

$$S_N(f) = \int_{-\infty}^{\infty} G(t)\,e^{-i2\pi ft}\,dt. \tag{3.57}$$

The inverse relationship also holds

$$G(t) = \int_{-\infty}^{\infty} S_N(f)\,e^{i2\pi ft}\,df. \tag{3.58}$$

Through the properties of the Fourier transform, the delta function associated with zero delay in Eq. (3.50) becomes the asymptotic value for large frequencies,

$$\lim_{f\to\infty} S_N(f) = \mathrm{E}[\mu] = 1/\mathrm{E}[\tau]. \tag{3.59}$$

The expression $S_N(f)\,/\mathrm{E}[\mu]$ therefore provides a normalized form of this spectrum. The large-delay asymptotic value for the coincidence rate in Eq. (3.51) becomes a delta function at zero frequency,

$$\lim_{\substack{\epsilon>0\\ \epsilon\to 0}} \int_{-\epsilon}^{\epsilon} S_N(f)\,df = \lim_{\substack{\epsilon>0\\ \epsilon\to 0}} \int_{-\epsilon}^{\epsilon} \delta(f)\,\mathrm{E}^2[\mu]\,df = \mathrm{E}^2[\mu]. \tag{3.60}$$

[8] We generally compute the point-process spectrum $S_N(f)$ for a theoretical construct but make use of the rate spectrum $S_\lambda(f,T)$ for actual data (see Sec. 3.4.5); the two spectra thus represent probabilistic and statistical measures, respectively.

This delta function also appears in $S_\lambda(f,T)$ and, with a different prefactor, in $S_\tau(\mathrm{f})$.

Combining the results obtained earlier yields (Lowen & Teich, 1993a; Lowen, 1996)

$$F(T) = \frac{2}{\pi^2 \,\mathrm{E}[\mu]\, T} \int_{0+}^{\infty} S_N(f) \sin^2(\pi f T)\, f^{-2}\, df \tag{3.61}$$

$$A(T) = \frac{4}{\pi^2 \,\mathrm{E}[\mu]\, T} \int_{0+}^{\infty} S_N(f) \sin^4(\pi f T)\, f^{-2}\, df, \tag{3.62}$$

as shown in Probs. 3.11 and 3.12, where the notation $0+$ indicates that the integral does not include the delta function at zero frequency. We can make use of Eq. (3.61) to obtain formulas for $F(T)$ and $A(T)$ in the limit of large counting times. Using the substitution $x \equiv \pi f T$, this equation becomes

$$F(T) = \frac{2}{\pi \,\mathrm{E}[\mu]} \int_{0+}^{\infty} S_N(x/\pi T) \sin^2(x)\, x^{-2}\, dx. \tag{3.63}$$

In the limit of large counting times, the spectrum approaches its low-frequency limit, whereupon Eq. (3.63) yields

$$\begin{aligned} \lim_{T\to\infty} F(T) &= \frac{2}{\pi \,\mathrm{E}[\mu]} \int_{0+}^{\infty} \lim_{f\to 0} S_N(f) \sin^2(x)\, x^{-2}\, dx \\ &= \frac{2}{\pi \,\mathrm{E}[\mu]} \lim_{f\to 0} S_N(f) \int_{0+}^{\infty} \sin^2(x)\, x^{-2}\, dx \\ &= \mathrm{E}[\tau] \lim_{f\to 0} S_N(f). \end{aligned} \tag{3.64}$$

Substituting this limit into Eq. (3.41) gives the same result for the normalized Haar-wavelet variance, $A(T)$.

In the opposite limit of small counting times and large frequencies we have

$$\lim_{T\to 0} F(T) = \mathrm{E}[\tau] \lim_{f\to\infty} S_N(f) = 1, \tag{3.65}$$

so that plots of the normalized Haar-wavelet variance and normalized point-process spectrum appear to be mirror images of each other.

In contrast to the coincidence rate, the estimation of $S_N(f)$ from a data set proves straightforward. A simple method (without averaging) follows from Eq. (3.49) and the Weiner–Khintchine theorem,

$$S_N(f) = \frac{1}{L}\mathrm{E}\left[\left|\sum_k e^{-i2\pi f t_k}\right|^2\right], \tag{3.66}$$

where $\{t_k\}$ again represents the set of times at which the events occur (rather than the times between events) and L denotes the duration of the data set. This method suffers from a major drawback: the times $\{t_k\}$ can take any of a continuous range

of values, which precludes the use of the fast Fourier-transform algorithm. However, the rate spectrum set forth in Sec. 3.4.5 proves amenable to this transform, yielding an efficient and practical method for estimating the spectrum of a point process.

A direct relationship exists between the rate spectrum and the point-process spectrum,

$$S_\lambda(f,T) = \sum_{k=-\infty}^{\infty} S_N(f + k/T) \frac{\sin^2(\pi f T)}{(\pi f T + \pi k)^2}\,, \tag{3.67}$$

and the two quantities differ only slightly for $T \ll 1/f$ (see Prob. 3.13). For the accurate estimation of $S_N(f)$ from a given sample of a point process with duration L, we choose an integer n such that $2^{n-1}/L$ exceeds the greatest frequency of interest by a large margin, then set $T = L/2^n$ and use the fast Fourier transform to obtain frequency-domain values from the 2^n elements of $\{Z_k(T)\}$.

3.5.3 General-wavelet variance

In Sec. 3.4.3 we made use of the Haar wavelet basis (Haar, 1910) in defining the normalized wavelet variance, because the form of the Haar wavelet leads to a simple representation in terms of the counting process $\{Z_k(T)\}$. Extensions to a general wavelet basis prove possible, although at the expense of more complex computations.

We again consider the variance of the continuous wavelet transform provided in Eq. (3.35), and employ its zero-mean property to obtain (Teich et al., 1996)

$$\begin{aligned}
\mathrm{Var}[C_{\psi,N}(a,b)] &= \mathrm{E}[C^2_{\psi,N}(a,b)] \\
&= \mathrm{E}\left[\int_s \int_t a^{-1}\,\psi[(s-b)/a]\,\psi[(t-b)/a]\,dN(s)\,dN(t)\right] \\
&= a^{-1}\int_s \int_t \psi[(s-b)/a]\,\psi[(t-b)/a]\,G(s-t)\,ds\,dt \\
&= a\int_x G(ax)\int_y \psi(x+y)\,\psi(y)\,dy\,dx \qquad (3.68) \\
&= a\int G(ax)\,C_{\psi,\psi}(1,x)\,dx \\
&= \int G(t)\,C_{\psi,\psi}(1,t/a)\,dt, \qquad (3.69)
\end{aligned}$$

where $C_{\psi,\psi}(1,x)$, the continuous wavelet transform of the wavelet function itself, does not depend on the coincidence rate $G(t)$ (Teich et al., 1996). As with the spectrum of the point process $S_N(f)$, the continuous nature of the times $\{t_k\}$ precludes the use of fast transform algorithms. Again, approximating the point process $dN(t)$ with the sequence of counts $\{Z_k(T)\}$ proves useful.

3.5.4 Generalized dimension

The **generalized dimension** D_q, which is closely related to the **Rényi entropy** (Rényi, 1955, 1970; Theiler, 1990), extends the simple concept of dimension developed in

Sec. 2.1 and provides a direct measurement of the fractal properties of an object. In the context of a collection of points, we consider a point process $dN(t)$ over the range of times $0 \leq t \leq L$ and employ the sequence of counts $\{Z_k\}$ to obtain

$$D_q \equiv \frac{1}{q-1} \lim_{T \to 0} \frac{\mathrm{E}\left\{\log\left[\sum_k Z_k^q(T)\right]\right\}}{\log(T)}, \tag{3.70}$$

where the sum extends over all non-empty counts. Note that for a given value of T, k assumes a maximum value of L/T.[9]

We can calculate the generalized dimension for any real value of q. Several values of q correspond to well-known generalized dimensions (Mandelbrot, 1982, Chapter 39), such as the **capacity dimension** D_0 (Pontrjagin & Schnirelmann, 1932) first discussed in Sec. 2.1.1, which is also called the **box-counting dimension**; the **information dimension** $\lim_{q \to 1} D_q$, which is closely related to the Kolmogorov entropy; and the **correlation dimension** D_2 (Grassberger & Procaccia, 1983). We also consider D_{-1} and $D_{1/2}$ (see Prob. 5.5). As reported in Sec. 2.1, the values of D_q for any object must lie between the **topological dimension** of the object and the **Euclidean dimension** of the space in which the object resides: zero for a point, unity for a line segment, two for a square, and so on.

Points on a line, as derived from a realization of a point process for example, will exhibit values of D_q between zero (the dimension of a point) and unity (the dimension of a line). In general, D_q for a nonfractal object assumes the lower bound given by the topological dimension of the object, for all indices q; the quantity $D_q = D$ is then an integer. For a (mono)fractal object, again $D_q = D$ for all q, but in contrast to the nonfractal object, D is not integer; indeed this forms one definition of a fractal (see Sec. 2.3). Finally, for a multifractal, D_q monotonically decreases with increasing q (Theiler, 1990).

Wavelet-based methods for estimating D_q also exist (Argoul, Arneodo, Elezgaray & Grasseau, 1989; Bacry, Muzy & Arneodo, 1993; Arrault & Arneodo, 1997); these can provide localized values of the generalized dimension.

We make special mention of a fractal dimension that does not belong to the D_q family of generalized dimensions: the **Hausdorff–Besicovitch dimension** D_{HB} (Mandelbrot, 1982, pp. 362–365 and references therein). For many simple cases in which $D_q = D$ for all q (such as the Cantor set), it turns out that $D_{\mathrm{HB}} = D$. Calculating this dimension involves fewer assumptions than determining D_q; the axes of the space within which the object exists need not be specified, nor need the dimension of that embedding space. However, determining D_{HB} proves far more difficult than calculating D_q for analytic examples, and it presents significant difficulties when applied to data. We therefore deal little with the Hausdorff–Besicovitch dimension in this book.

The practical use of the generalized dimension D_q requires modification when applied to real point processes. For any orderly point process, the number of events

[9] If we set $q = 0$ and make the identifications $T = \epsilon$ and $Z(T) = M(\epsilon)$, we recover Eq. (2.2).

$N(L)$ occurring between the origin and a maximum time L assumes a finite value with probability one. Therefore, a minimum interevent time $\tau_{\min}$ exists; for $T < \tau_{\min}$, for all k either $Z_k(T) = 0$ or $Z_k(T) = 1$. Equation (3.70) then becomes

$$\begin{aligned} D_q &= \frac{1}{q-1} \lim_{T\to 0} \frac{\mathrm{E}\Big\{\log\Big[\sum_k 1^q\Big]\Big\}}{\log(T)} \\ &= \frac{1}{q-1} \lim_{T\to 0} \frac{\mathrm{E}\big\{\log\left[N(L)\right]\big\}}{\log(T)} \\ &= \frac{\mathrm{E}\big\{\log\left[N(L)\right]\big\}}{q-1} \lim_{T\to 0} \frac{1}{\log(T)} \\ &= 0, \end{aligned} \tag{3.71}$$

where the sum again does not contain empty counts. Hence, for a finite collection of points, we obtain $D_q = 0$, a result not indicative of fractal characteristics.

In the context of point processes, we therefore modify the definition of the generalized dimension to accommodate scaling behavior over a range of counting times T. We recast Eq. (3.70) in terms of a **scaling equation**,

$$\sum_k Z_k^q(T) \sim T^{(q-1)Dq}\,, \tag{3.72}$$

which we can alternatively write in the form of a **generalized-dimension scaling function**:

$$\eta_q(T) \equiv \Big[\sum_k Z_k^q(T)/N(L)\Big]^{\frac{1}{q-1}} \approx T^{Dq}\,, \tag{3.73}$$

normalized such that $\eta_q(T \to 0) = 1$. If Eqs. (3.72) and (3.73) hold over a range of counting times T, then the resulting exponents on the right-hand sides of these equations yield the generalized dimensions D_q. This directly reveals the fractal properties of point-process sample paths. In practice, the D_q assume noninteger values only for the class of **fractal point processes**, and not for the class of **fractal-rate point processes** that largely form the focus of this book (see Sec. 5.5 and Prob. 5.5).

In the special case when $q = 0$, Eq. (3.73) reduces to

$$\eta_0(T) \equiv \frac{N(L)}{\sum_k Z_k^0(T)} \approx T^{D0}\,, \tag{3.74}$$

which yields $D_0 = 0$ for $T < \tau_{\min}$, in accordance with Eq. (3.71), since the sum in the denominator of Eq. (3.74) is then $N(L)$.[10]

[10] This power-law dependence over a range of counting times recalls Eq. (1.1) rather than Eq. (2.1).

3.5.5 Correlation measures for pairs of point processes

Second-order methods prove useful for revealing correlations between sequences of events, which indicate how information is shared between pairs of point processes.[11] Although such methods may not detect all of the subtle forms of interdependence to which information-theoretic approaches are sensitive (see, for example, Kabanov, 1978; Rieke, Warland, de Ruyter van Steveninck & Bialek, 1997; Dayan & Abbott, 2001), the latter methods suffer from limitations arising from the finite sizes of many real data sets (see Lowen, Ozaki, Kaplan & Teich, 1998).

We consider two second-order measures, in turn: the **normalized Haar-wavelet covariance** and the **cross-spectrum**. They exhibit different immunity to nonstationarities, and different tradeoffs between bias and variance, just as their single-dimensional counterparts do (see Chapter 12). Both measures prove useful in the analysis of pairs of point processes.[12]

- *Normalized Haar-wavelet covariance.* We define the normalized Haar-wavelet covariance $A^{(2)}(T)$ as a generalization of the normalized Haar-wavelet variance $A(T)$ defined in Sec. 3.4.3. This measure is insensitive to constant values and can be rendered insensitive to higher-order polynomial trends by making use of other wavelets (see Sec. 3.5.3). We initially developed this measure to analyze correlations between pairs of visual-system spike trains, such as those recorded from retinal ganglion cells and cells in the lateral geniculate nucleus (Lowen et al., 2001). At first, we referred to $A^{(2)}(T)$ as the "normalized wavelet cross-correlation function," but we prefer the designation "normalized Haar-wavelet covariance" since it highlights the relationship between $A^{(2)}(T)$ and $A(T)$.

  The computation of the normalized Haar-wavelet covariance at a particular counting time T begins with the division of both point processes into contiguous counting durations T. For the first point process, we register the number of events $Z_{1,k}$ that fall within the kth duration for all indices k. Next we compute the difference between the count numbers in a given duration, $Z_{1,k}$, and the duration that immediately follows, $Z_{1,k+1}$, for all k, much as when computing the normalized Haar-wavelet variance. We then carry out the same procedure for the second point process, beginning with the number of events $Z_{2,k}$ that fall within the kth duration.

[11] For example, it may be of interest to study how information is shared between the point processes at the input and output of a cell. Such a pair of processes collectively forms a **bivariate point process**, a form of marked point process (see Cox & Isham, 1980, Chapter 5).

[12] They have been used to reveal unexpected correlations between pairs of visual-system spike trains (Lowen et al., 2001) and between earthquakes and geoelectrical extreme events (Telesca, Balasco, Colangelo, Lapenna & Macchiato, 2004), as examples.

In analogy with the definition of the normalized Haar-wavelet variance, we define the normalized Haar-wavelet covariance as:

$$A^{(2)}(T) \equiv \frac{\mathrm{E}\{[Z_{1,k}(T) - Z_{1,k+1}(T)]\,[Z_{2,k}(T) - Z_{2,k+1}(T)]\}}{2\{\mathrm{E}[Z_{1,k}(T)]\,\mathrm{E}[Z_{2,k}(T)]\}^{1/2}}\,. \quad (3.75)$$

The normalization imposed in Eq. (3.75) gives rise to three salutary features for $A^{(2)}(T)$: (1) it is symmetric in the two point processes; (2) it reduces to the marginal normalized wavelet variance $A(T)$ if the two point processes are identical — in particular, it assumes a value of unity for all counting times T if both point processes comprise the same homogeneous Poisson process, in analogy with the normalized wavelet variance $A(T)$.

- *Cross-spectrum.* The cross-spectrum $S_N^{(2)}(f)$ is a generalization of the point-process spectrum for individual spike trains, in much the same way as the normalized Haar-wavelet covariance derives from the normalized Haar-wavelet variance (Lowen et al., 2001). Although not often used, this measure has a long history in the annals of statistics. It appears to have been first introduced by Jenkins (1961) and its use has been advanced by Brillinger (1986).

  A number of definitions for the cross-spectrum have been put forward; we choose one that is real and symmetric in the two point processes, and reduces to the single-process version when the two processes coincide. An extension of Eq. (3.66) leads to the cross-spectrum

$$S_N^{(2)}(f) \equiv \frac{1}{L}\mathrm{E}\left[\mathrm{Re}\left\{\sum_k e^{-i2\pi f t_{1,k}} \sum_m e^{i2\pi f t_{2,m}}\right\}\right], \quad (3.76)$$

  where $\{t_{1,k}\}$ and $\{t_{2,k}\}$ index the events in point processes 1 and 2, respectively, occurring in a time of duration L. Equation (3.76) is indeed symmetric, and reduces to Eq. (3.66) for identical sets $\{t_{1,k}\}$ and $\{t_{2,k}\}$. Although it returns a real result, the cross-spectrum can assume negative values. For example, two Poisson processes modulated by sinewaves of identical frequency, but opposite phase, yield $S_N^{(2)}(f) < 0$ near the modulation frequency and its harmonics. For independent spike trains, we have $S_N^{(2)}(f) = 0$.

  As with the single-process spectrum, for the purposes of practical estimation it proves easier to employ the rate-based version

$$S_\lambda^{(2)}(f,T) \equiv \frac{1}{L}\mathrm{E}\left[\mathrm{Re}\left\{\sum_k Z_{1,n+k}(T) e^{-i2\pi k f T} \sum_m Z_{2,n+m}(T) e^{i2\pi m f T}\right\}\right], \quad (3.77)$$

  where $\{Z_{1,n}(T)\}$ and $\{Z_{2,n}(T)\}$ describe the counts for the two point processes.

Problems

3.1 *Point-process models* For each of the following examples, specify whether an orderly, one-dimensional point process provides a useful model. For the remainder, describe a modification of the example that would make the model apply.

1. longitude and latitude of trees on Long Island, New York;
2. the times of raindrops hitting a roof;
3. the arrival times of customers at an automatic teller machine;
4. thunderstorm occurrence times in San Diego county during February 1993;
5. the times at which cars are in the Ted Williams tunnel, which passes under Boston harbor;
6. the set of numbers $1/(n + x_n)$, where n ranges over all positive integers and $\{x_n\}$ is a set of independent exponentially distributed random variables of unit mean;
7. the times of maximum daily temperature at the summit of Mount Everest;
8. all human heartbeat times (defined as the time of maximum contraction) from anyone anywhere on the planet, as transmitted to a central recording station;
9. two random numbers selected uniformly from the unit interval;
10. the sign of the difference between the Dow Jones Industrial Average and the previous day's closing price.

3.2 *Short-time normalized variance for an orderly point process* Justify the approximation $\mathrm{E}\left[Z^2(T)\right] \approx \mathrm{E}[Z(T)]$ in Eq. (3.34).

3.3 *Connection between interval- and count-based statistics* Using Eqs. (3.11) and (3.29), show that Eq. (3.30) holds.

3.4 *Forward-recurrence-time distribution* Explain why Eq. (3.29) holds.

3.5 *Skewness and kurtosis values* As defined in Eq. (3.4), what values can the skewness and kurtosis assume? How does prohibiting negative interevent intervals change this?

3.6 *Infinite moments of the Lévy density* Show that the random variable corresponding to the probability density function given by Eq. (3.13) has infinite moments for all positive integer orders. For which fractional orders do its moments exist?

3.7 *Rescaled range and detrended fluctuations for independent intervals* Provide a heuristic argument showing that the rescaled range statistic $U(k)$ and detrended fluctuation analysis $Y(k)$ indeed vary as $\sqrt{k}$ for large k and independent intervals with finite variance.

3.8 *Interval- and forward-recurrence-time statistics* Show that Eq. (3.11) is valid.

3.9 *Interval and point-process spectra* Show that the interval-based spectrum $S_\tau(\mathrm{f})$ and the point-process spectrum $S_N(f)$ are proportional to each other, at low frequencies, for interevent intervals with $\mathrm{Var}[\tau]/\mathrm{E}^2[\tau] \ll 1$.

3.10 *Normalized variance and coincidence rate* Prove Eq. (3.52).

3.11 *Normalized variance and point-process spectrum* Prove Eq. (3.61).

3.12 *Normalized Haar-wavelet variance and point-process spectrum* Show that Eq. (3.62) is valid.

3.13 *Connection between rate and point-process spectra* Prove Eq. (3.67), and show that the two spectra approach each other for small values of the product fT. Problem 4.8 explicitly illustrates this connection for the homogeneous Poisson and gamma renewal point processes.