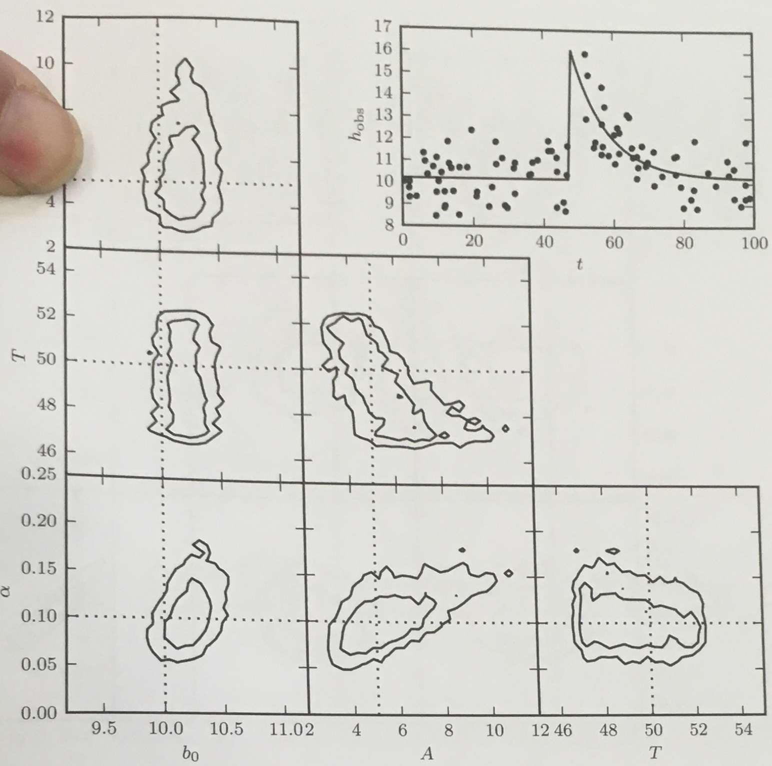

Figure 10.25. A matched filter search for a burst signal in time series data. A simulated data set generated from a model of the form $y(t) = b_0$ for $t < T$ and $y = b_0 + A\exp[-\alpha(t - T)]$ for $t > T$, with homoscedastic Gaussian errors with $\sigma = 1$, is shown in the top-right panel. The posterior pdf for the four model parameters is determined using MCMC and shown in the other panels.

not only in time, but in frequency as well, the wavelet-based analysis discussed in §10.2.4 is a good choice. A simple example of such an analysis is shown in figure 10.28. The two-dimensional wavelet-based PSD easily recovers the increase of characteristic chirp frequency with time. To learn more about such types of analysis, we refer the reader to the rapidly growing body of tools and publications developed in the context of gravitational wave analysis.[6]

10.5. Analysis of Stochastic Processes

Stochastic variability includes behavior that is not predictable forever as in the periodic case, but unlike temporally localized events, variability is always there. Typically, the underlying physics is so complex that we cannot deterministically predict future

[6] See, for example, http://www.ligo.org/

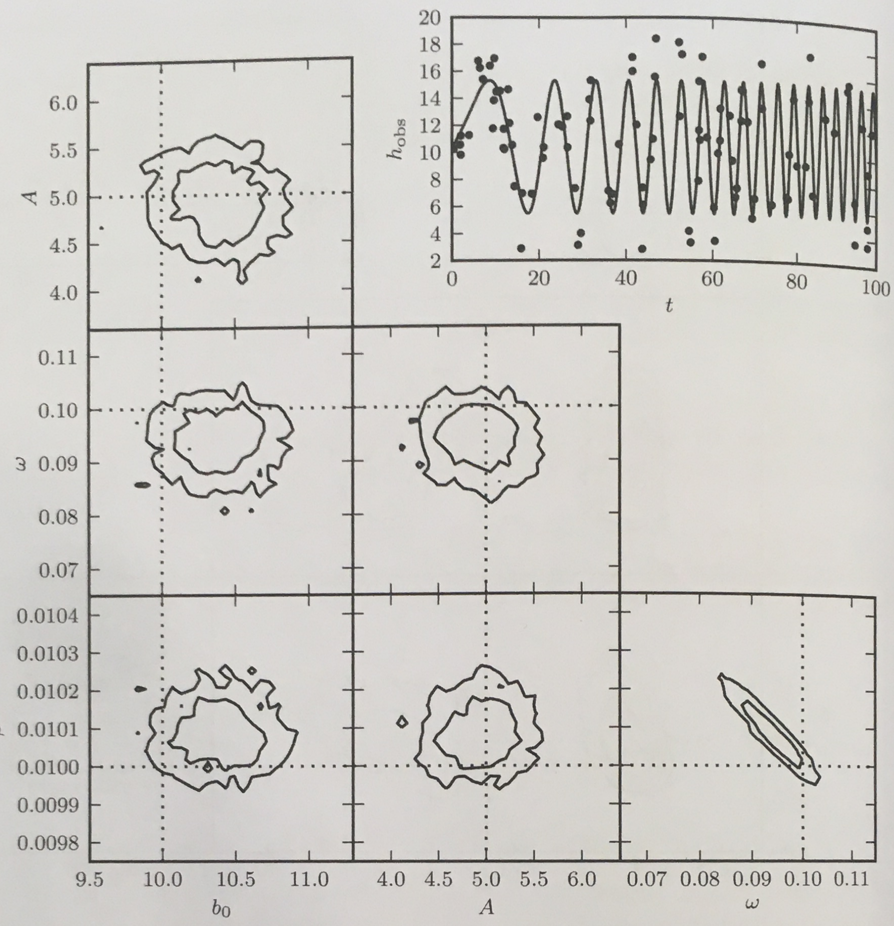

Figure 10.26. A matched filter search for a chirp signal in time series data. A simulated data set generated from a model of the form $y = b_0 + A\sin[\omega t + \beta t^2]$, with homoscedastic Gaussian errors with $\sigma = 2$, is shown in the top-right panel. The posterior pdf for the four model parameters is determined using MCMC and shown in the other panels.

values (i.e., the stochasticity is inherent in the process, rather than due to measurement noise). Despite their seemingly irregular behavior, stochastic processes can be quantified, too, as briefly discussed in this section. References to more in-depth literature on stochastic processes are listed in the final section.

10.5.1. The Autocorrelation and Structure Functions

One of the main statistical tools for the analysis of stochastic variability is the autocorrelation function. It represents a specialized case of the correlation function of two functions, $f(t)$ and $g(t)$, scaled by their standard deviations, and defined at time lag Δt as

$$\mathrm{CF}(\Delta t) = \frac{\lim_{T\to\infty} \frac{1}{T}\int_{(T)} f(t)\, g(t+\Delta t)\, dt}{\sigma_f\, \sigma_g}, \tag{10.89}$$

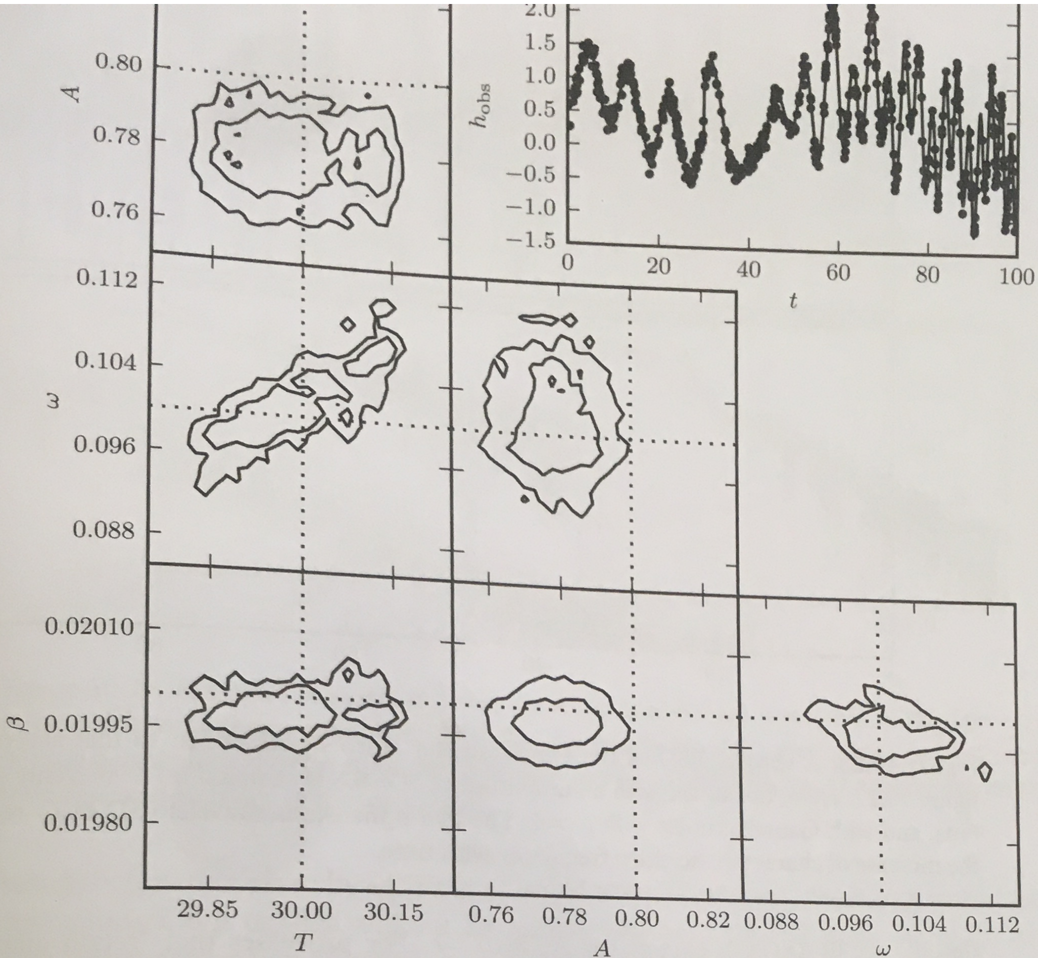

Figure 10.27. A ten-parameter chirp model (see eq. 10.87) fit to a time series. Seven of the parameters can be considered nuisance parameters, and we marginalize over them in the likelihood contours shown here.

where σ_f and σ_g are standard deviations of $f(t)$ and $g(t)$, respectively. With this normalization, the correlation function is unity for $\Delta t = 0$ (without normalization by standard deviation, the above expression is equal to the covariance function). It is assumed that both f and g are statistically weakly stationary functions, which means that their mean and autocorrelation function (see below) do not depend on time (i.e., they are statistically the same irrespective of the time interval over which they are evaluated). The correlation function yields information about the time delay between two processes. If one time series is produced from another one by simply shifting the time axis by t_{lag}, their correlation function has a peak at $\Delta t = t_{\mathrm{lag}}$.

With $f(t) = g(t) = y(t)$, the autocorrelation of $y(t)$ defined at time lag Δt is

$$\mathrm{ACF}(\Delta t) = \frac{\lim_{T\to\infty} \frac{1}{T} \int_{(T)} y(t)\, y(t+\Delta t)\, dt}{\sigma_y^2}. \tag{10.90}$$

The autocorrelation function yields information about the variable timescales present in a process. When y values are uncorrelated (e.g., due to white noise without any

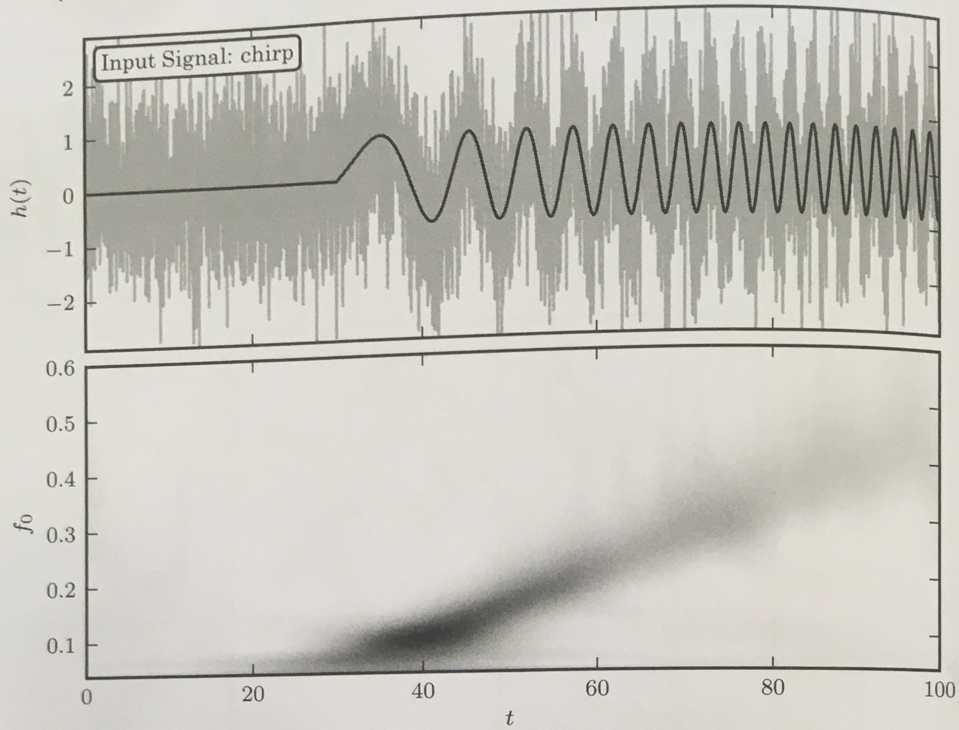

Figure 10.28. A wavelet PSD of the ten-parameter chirp signal similar to that analyzed in figure 10.27. Here, the signal with an amplitude of $A = 0.8$ is sampled in 4096 evenly spaced bins, and with Gaussian noise with $\sigma = 1$. The two-dimensional wavelet PSD easily recovers the increase of characteristic chirp frequency with time.

signal), $\mathrm{ACF}(\Delta t) = 0$, except for $\mathrm{ACF}(0) = 1$. For processes that "retain memory" of previous states only for some characteristic time τ, the autocorrelation function vanishes for $\Delta t \gg \tau$. In other words, the predictability of future behavior for such a process is limited to times up to $\sim\tau$. One such process is *damped random walk*, discussed in more detail in §10.5.4.

The autocorrelation function and the PSD of function $y(t)$ (see eq. 10.6) are Fourier pairs; this fact is known as the Wiener–Khinchin theorem and applies to stationary random processes. The former represents an analysis method in the time domain, and the latter in the frequency domain. For example, for a periodic process with a period P, the autocorrelation function oscillates with the same period, while for processes that retain memory of previous states for some characteristic time τ, ACF drops to zero for $t \sim \tau$.

The structure function is another quantity closely related to the autocorrelation function,

$$\mathrm{SF}(\Delta t) = \mathrm{SF}_\infty \, [1 - \mathrm{ACF}(\Delta t)]^{1/2}, \tag{10.91}$$

where SF_∞ is the standard deviation of the time series evaluated over an infinitely large time interval (or at least much longer than any characteristic timescale τ). The structure function, as defined by eq. 10.91, is equal to the standard deviation of the distribution of the difference of $y(t_2) - y(t_1)$ evaluated at many different t_1 and t_2 such that time lag $\Delta t = t_2 - t_1$, and divided by $\sqrt{2}$ (because of differencing). When the structure function $\mathrm{SF} \propto t^\alpha$, then $\mathrm{PSD} \propto 1/f^{(1+2\alpha)}$. In the statistics literature, the

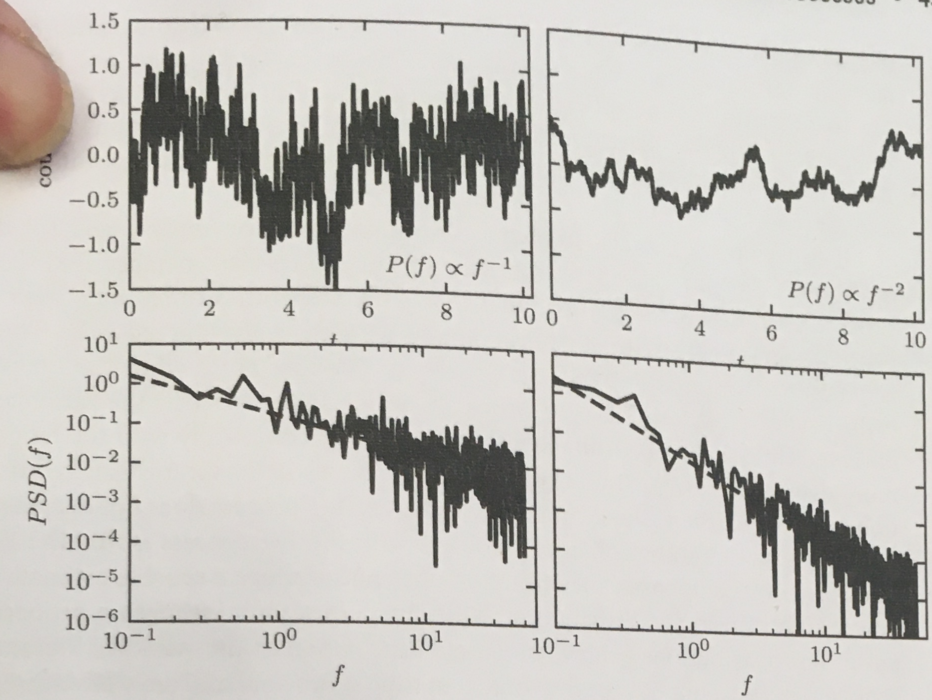

Figure 10.29. Examples of stochastic time series generated from power-law PSDs (left: $1/f$; right: $1/f^2$) using the method from [56]. The top panels show the generated data, while the bottom panels show the corresponding PSD (dashed lines: input PSD; solid lines: determined from time series shown in the top panels).

structure function given by eq. 10.91 is called the *second-order structure function* (or *variogram*) and is defined without the square root (e.g., see FB2012). Although the early use in astronomy followed the statistics literature, for example, [52], we follow here the convention used in recent studies of quasar variability, for example, [58] and [16] (the appeal of taking the square root is that SF then has the same unit as the measured quantity). Note, however, that definitions in the astronomical literature are not consistent regarding the $\sqrt{2}$ factor discussed above.

Therefore, a stochastic time series can be analyzed using the autocorrelation function, the PSD, or the structure function. They can reveal the statistical properties of the underlying process, and distinguish processes such as white noise, random walk (see below), and damped random walk (discussed in §10.5.4). They are mathematically equivalent and all are used in practice; however, due to issues of noise and sampling, they may not always result in equivalent inferences about the data.

Examples of Stochastic Processes: 1/f and 1/f² Processes

For a given autocorrelation function or PSD, the corresponding time series can be generated using the algorithm described in [56]. Essentially, the amplitude of the Fourier transform is given by the PSD, and phases are assigned randomly; the inverse Fourier transform then generates time series.

The connection between the PSD and the appearance of time series is illustrated in figure 10.29 for two power-law PSDs: $1/f$ and $1/f^2$. The PSD normalization is such that both cases have similar power at low frequencies. For this reason, the overall amplitudes (more precisely, the variances) of the two time series are similar. The power at high frequencies is much larger for the $1/f$ case, and this is why the corresponding time series has the appearance of noisy data (the top-left panel in figure 10.29). The structure function for the $1/f$ process is constant, and proportional to $t^{1/2}$ for the $1/f^2$ process (remember that we defined structure function with a square root).

The $1/f^2$ process is also known as *Brownian motion* or *random walk* (or *drunkard's walk*). For an exellent introduction from a physicist's perspective, see [26]. Processes whose PSD is proportional to $1/f$ are sometimes called *long-term memory processes* (mostly in the statistical literature), *flicker noise* and *red noise*. The latter is not unique as sometimes the $1/f^2$ process is called *red noise*, while the $1/f$ process is then called *pink noise*. The $1/f$ processes have infinite variance; the variance of an observed time series of a finite length increases logarithmically with the length (for more details, see [42]). Similarly to the behavior of the mean for Cauchy distribution (see §5.6.3), the variance of the mean for the $1/f$ process does not decrease with the sample size. Another practical problem with the $1/f$ process is that the Fourier transform of its autocovariance function does not produce a reliable estimate of the power spectrum in the distribution's tail. Difficulties with estimating properties of power-law distributions (known as *Pareto distribution* in the statistics literature) in general cases (i.e., not only in the context of time series analysis) are well summarized in [10].

AstroML includes a routine which generates power-law light curves based on the method of [56]. It can be used as follows:

```
>>> import numpy as np
>>> from astroML.time_series import generate_power_law

# beta gives the power-law index: P ~ f^-beta
>>> y = generate_power_law(N=1024, dt=0.01, beta=2)
```

This routine is used to generate the data shown in figure 10.29.

10.5.2. Autocorrelation and Structure Function for Evenly and Unevenly Sampled Data

In the case of evenly sampled data, with $t_i = (i-1)\Delta t$, the autocorrelation function of a discretely sampled $y(t)$ is defined as

$$\mathrm{ACF}(j) = \frac{\sum_{i=1}^{N-j}\left[(y_i - \bar{y})\,(y_{i+j} - \bar{y})\right]}{\sum_{i=1}^{N}(y_i - \bar{y})^2}. \tag{10.92}$$

With this normalization the autocorrelation function is dimensionless and ACF(0) = 1. The normalization by variance is sometimes skipped (see [46]), in which case a more appropriate name is the *covariance function*.

When a time series has a nonvanishing ACF, the uncertainty of its mean is larger than for an uncorrelated data set (cf. eq. 3.34),

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{N}} \left[1 + 2 \sum_{j=1}^{N} \left(1 - \frac{j}{N} \right) \mathrm{ACF}(j) \right]^{1/2}, \tag{10.93}$$

where σ is the homoscedastic measurement error. This fact is often unjustifiably neglected in analysis of astronomical data.

When data are unevenly sampled, the ACF cannot be computed using eq. 10.92. For the case of unevenly sampled data, Edelson and Krolik [22] proposed the *discrete correlation function* (DCF) in an astronomical context (called the *slot autocorrelation function* in physics). For discrete unevenly sampled data with homoscedastic errors, they defined a quantity

$$\mathrm{UDCF}_{ij} = \frac{(y_i - \overline{y})\,(g_j - \overline{g})}{\left[(\sigma_y^2 - e_y^2)\,(\sigma_g^2 - e_g^2) \right]^{1/2}}, \tag{10.94}$$

where e_y and e_g are homoscedastic measurement errors for time series y and g. The associated time lag is $\Delta t_{ij} = t_i - t_j$. The discrete correlation function at time lag Δt is then computed by binning and averaging UDCF_{ij} over M pairs of points for which $\Delta t - \delta t/2 \le \Delta t_{ij} \le \Delta t + \delta t/2$, where δt is the bin size. The bin size is a trade-off between accuracy of $\mathrm{DCF}(\Delta t)$ and its resolution. Edelson and Krolik showed that even uncorrelated time series will produce values of the cross-correlation $\mathrm{DCF}(\Delta t) \sim \pm 1/\sqrt{M}$.

With its binning, this method is similar to procedures for computing the structure function used in studies of quasar variability [15, 52]. The main downside of the DCF method is the assumption of homoscedastic error. Nevertheless, heteroscedastic errors can be easily incorporated by first computing the structure function, and then obtaining the ACF using eq. 10.91. The structure function is equal to the intrinsic distribution width divided by $\sqrt{2}$ for a bin of Δt_{ij} (just as when computing the DCF above). This width can be estimated for heteroscedastic data using eq. 5.69, or the corresponding exact solution given by eq. 5.64.

Scargle has developed different techniques to evaluate the discrete Fourier transform, correlation function and autocorrelation function of unevenly sampled time series (see [46]). In particular, the discrete Fourier transform for unevenly sampled data and the Wiener–Khinchin theorem are used to estimate the autocorrelation function. His method also includes a prescription for correcting the effects of uneven sampling, which results in leakage of power to nearby frequencies (the so-called *side-lobe effect*). Given an unevenly sampled time series, $y(t)$, the essential steps of Scargle's procedure are as follows:

1. Compute the generalized Lomb–Scargle periodogram for $y(t_i)$, $i = 1, \ldots, N$, namely $P_{\mathrm{LS}}(\omega)$.
2. Compute the sampling window function using the generalized Lomb–Scargle periodogram using $z(t_i) = 1$, $i = 1, \ldots, N$, namely $P_{\mathrm{LS}}^{W}(\omega)$.
3. Compute inverse Fourier transforms for $P_{\mathrm{LS}}(\omega)$ and $P_{\mathrm{LS}}^{W}(\omega)$, namely $\rho(t)$ and $\rho^{W}(t)$, respectively.
4. The autocorrelation function at lag t is $\mathrm{ACF}(t) = \rho(t)/\rho^{W}(t)$.

AstroML includes tools for computing the ACF using both Scargle's method and the Edelson and Krolik method:

```
>>> import numpy as np
>>> from astroML.time_series import generate_damped_RW
>>> from astroML.time_series import ACF_scargle, ACF_EK

>>> t = np.arange(0, 1000)
>>> y = generate_damped_RW(t, tau=300)
>>> dy = 0.1
>>> y = np.random.normal(y, dy)

# Scargle's method
>>> ACF, bins = ACF_scargle(t, y, dy)

# Edelson-Krolik method
>>> ACF, ACF_err, bins = ACF_EK(t, y, dy)
```

For more detail, see the source code of figure 10.30.

Figure 10.30 illustrates the use of Edelson and Krolik's DCF method and the Scargle method. They produce similar results; errors are easier to compute for the DCF method and this advantage is crucial when fitting models to the autocorrelation function.

Another approach to estimating the autocorrelation function is direct modeling of the correlation matrix, as discussed in the next section.

10.5.3. Autoregressive Models

Autocorrelated time series can be analyzed and characterized using stochastic *autoregressive models*. Autoregressive models provide a good general description of processes that "retain memory" of previous states (but are not periodic). An example of such a model is the random walk, where each new value is obtained by adding noise to the preceding value:

$$y_i = y_{i-1} + e_i. \tag{10.95}$$

When y_{i-1} is multiplied by a constant factor greater than 1, the model is known as a *geometric random walk* model (used extensively to model stock market data). The noise need not be Gaussian; white noise consists of uncorrelated random variables with zero mean and constant variance, and Gaussian white noise represents the most common special case of white noise.

The random walk can be generalized to the *linear autoregressive* (linear AR) model with dependencies on k past values (i.e., not just one as in the case of random walk). An autoregressive process of order k, AR(k), for a discrete data set is defined by

$$y_i = \sum_{j=1}^{k} a_j\, y_{i-j} + e_i. \tag{10.96}$$

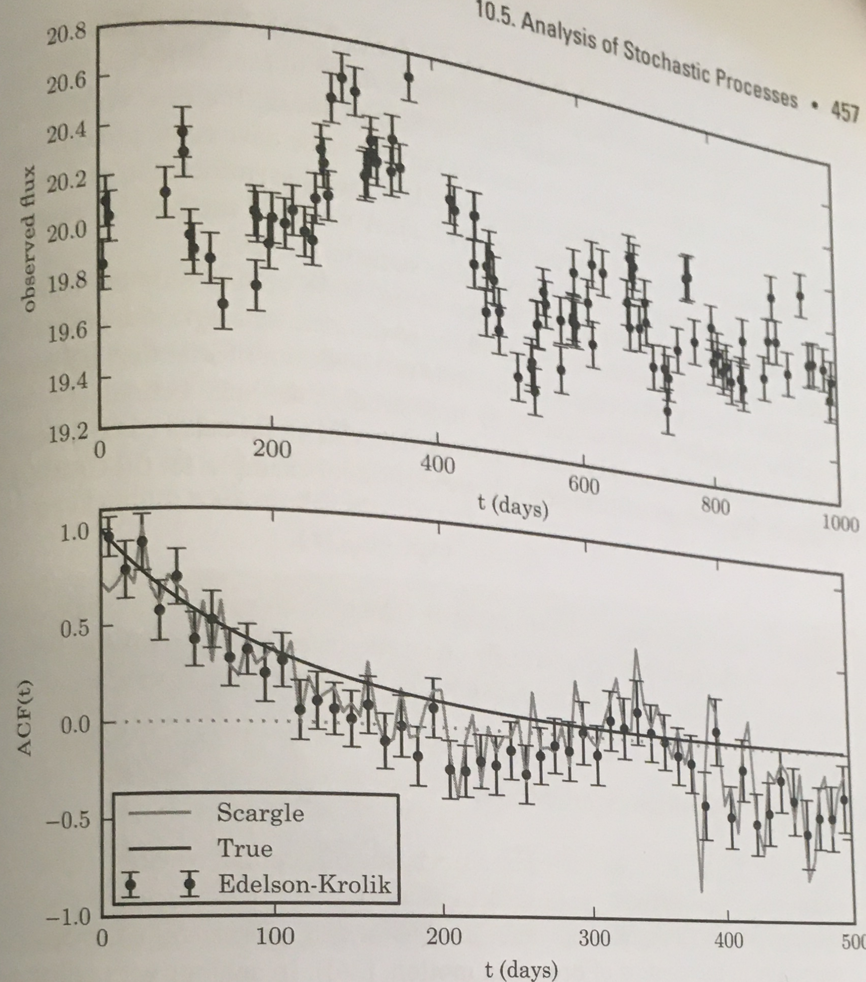

Figure 10.30. Example of the autocorrelation function for a stochastic process. The top panel shows a simulated light curve generated using a damped random walk model (§10.5.4). The bottom panel shows the corresponding autocorrelation function computed using Edelson and Krolik's DCF method and the Scargle method. The solid line shows the input autocorrelation function used to generate the light curve.

That is, the latest value of y is expressed as a linear combination of the k previous values of y, with the addition of noise (for random walk, $k = 1$ and $a_1 = 1$). If the data are drawn from a stationary process, coefficients a_j satisfy certain conditions. The ACF for an AR(k) process is nonzero for all lags, but it decays quickly.

The literature on autoregressive models is abundant because applications vary from signal processing and general engineering to stock-market modeling. Related modeling frameworks include the moving average (MA, where y_i depends only on past values of noise), autoregressive moving average (ARMA, a combination of AR and MA processes), autoregressive integrated moving average (ARIMA, a combination of ARMA and random walk), and state-space or dynamic linear modeling (so-called Kalman filtering). More details and references about these stochastic autoregressive models can be found in FB2012. Alternatively, modeling can be done in the frequency domain (per the Wiener–Khinchin theorem).

For example, a simple but astronomically very relevant problem is distinguishing a random walk from pure noise. That is, given a time series, the question is whether it better supports the hypothesis that $a_1 = 0$ (noise) or that $a_1 = 1$ (random walk). For comparison, in stock market analysis this pertains to predicting the next data value based on the current data value and the historic mean. If a time series is a random walk, values higher and lower than the current value have equal probabilities. However, if a time series is pure noise, there is a useful asymmetry in probabilities due to regression toward the mean (see §4.7.1). A standard method for answering this question is to compute the Dickey–Fuller statistic; see [20].

An autoregressive process defined by eq. 10.96 applies only to evenly sampled time series. A generalization is called the *continuous autoregressive process*, CAR(k); see [31]. The CAR(1) process has recently received a lot of attention in the context of quasar variability and is discussed in more detail in the next section.

In addition to autoregressive models, data can be modeled using the covariance matrix (e.g., using Gaussian process; see §8.10). For example, for the CAR(1) process,

$$S_{ij} = \sigma^2 \exp(-|t_{ij}|/\tau), \tag{10.97}$$

where σ and τ are model parameters; σ^2 controls the short timescale covariance ($t_{ij} \ll \tau$), which decays exponentially on a timescale given by τ. A number of other convenient models and parametrizations for the covariance matrix are discussed in the context of quasar variability in [64].

10.5.4. Damped Random Walk Model

The CAR(1) process is described by a stochastic differential equation which includes a damping term that pushes $y(t)$ back to its mean (see [31]); hence, it is also known as *damped random walk* (another often-used name is the Ornstein–Uhlenbeck process, especially in the context of Brownian motion, [26]). In analogy with calling random walk "drunkard's walk," damped random walk could be called "married drunkard's walk" (who always comes home instead of drifting away).

Following eq. 10.97, the autocorrelation function for a damped random walk is

$$\mathrm{ACF}(t) = \exp(-t/\tau), \tag{10.98}$$

where τ is the characteristic timescale (relaxation time, or damping timescale). Given the ACF, it is easy to show that the structure function is

$$\mathrm{SF}(t) = \mathrm{SF}_\infty \left[1 - \exp(-t/\tau)\right]^{1/2}, \tag{10.99}$$

where SF_∞ is the asymptotic value of the structure function (equal to $\sqrt{2}\sigma$, where σ is defined in eq. 10.97, when the structure function applies to differences of the analyzed process; for details see [31, 37]) and

$$\mathrm{PSD}(f) = \frac{\tau^2\, \mathrm{SF}_\infty^2}{1 + (2\pi f \tau)^2}. \tag{10.100}$$

Therefore, the damped random walk is a $1/f^2$ process at high frequencies, just as ordinary random walk. The damped nature is seen as the flat PSD at low frequencies ($f \ll 2\pi/\tau$). An example of a light curve generated using a damped random walk is shown in figure 10.30.

For evenly sampled data, the CAR(1) process is equivalent to the AR(1) process with $a_1 = \exp(-1/\tau)$; that is, the next value of y is the damping factor times the previous value plus noise. The noise for the AR(1) process, σ_{AR}, is related to SF_∞ via

$$\sigma_{AR} = \frac{SF_\infty}{\sqrt{2}} \left[1 - \exp(-2/\tau)\right]^{1/2}. \tag{10.101}$$

A damped random walk provides a good description of the optical continuum variability of quasars; see [31, 33, 37]. Indeed, this model is so successful that it has been used to distinguish quasars from stars (both are point sources in optical images, and can have similar colors) based solely on variability behavior; see [9, 36]. Nevertheless, at short timescales of the order a month or less (at high frequencies from 10^{-6} Hz up to 10^{-5} Hz), the PSD is closer to $1/f^3$ behavior than to $1/f^2$ predicted by the damped random walk model; see [39, 64].

Scikit-learn contains a utility which generates damped random walk light curves given a random seed:

```
>>> import numpy as np
>>> from astroML.time_series import generate_damped_RW

>>> t = np.arange(0, 1000)
>>> y = generate_damped_RW(t, tau=300, random_state=0)
```

For a more detailed example, see the source code associated with figure 10.30.

10.6. Which Method Should I Use for Time Series Analysis?

Despite extensive literature developed in the fields of signal processing, statistics, and econometrics, there are no universal methods that always work. This is even more so in astronomy where uneven sampling, low signal-to-noise ratio, and heteroscedastic errors often prevent the use of standard methods drawn from other fields.

The main tools for time series analysis belong to either the time domain or the frequency domain. When searching for periodic variability, tools from the frequency domain are usually better because the signal becomes more "concentrated." This is a general feature of model fitting, where a matched filter approach can greatly improve the ability to detect a signal. For typical astronomical periodic time series, the generalized Lomb–Scargle method is a powerful method; when implemented to model several terms in a truncated Fourier series, instead of a single sinusoid, it works well when analyzing variable stars. It is well suited to unevenly sampled data with low signal-to-noise ratio and heteroscedastic errors. Nevertheless, when the shape of the underlying light curve cannot be approximated with a small number of Fourier terms, nonparametric methods such as the minimum string length method, the phase dispersion minimization method, or the Bayesian blocks algorithm may perform better. Analysis of arrival time data represents different challenges; the Gregory and Loredo algorithm is a good general method in this case.