

Simulation of Time Series

11.0 Introduction

Of the many reasons for estimating a hypothetical spectral density function (SDF) for a time series $\{X_t\}$, a prominent one is to use the estimated SDF as the basis for creating computer-generated artificial series with characteristics similar to those of $\{X_t\}$. Simulated series are of interest in bootstrapping, a popular computer-based method for assessing the sampling variability in certain statistics (see, e.g., Efron and Gong, 1983, or Davison and Hinkley, 1997; we touch upon this topic in Sections 11.4 to 11.6). Simulated series are also useful for assessing procedures that act upon time series. As a concrete example, consider the atomic clock data $\{X_t\}$ displayed in Figure 326(a). These data compare the time kept by two atomic clocks, both of which were part of a large ensemble of atomic clocks maintained at the US Naval Observatory and used to provide a standard of time for the United States. Since no single clock keeps perfect time, pairwise comparisons amongst clocks are used to form a nonphysical “paper” clock that in principle keeps time better than any individual physical clock. Various procedures – known as time-scale algorithms – have been proposed for manipulating pairwise comparisons to form a paper clock (see, e.g., references in Matsakis and Tavella, 2008). The relative merits of different algorithms can be assessed by applying them to simulated pairwise comparisons. One way to ensure realistic simulated series is to make them consistent with SDF estimates similar to those in Figure 329.

In what follows we first formulate procedures for generating artificial time series that can be considered to be realizations from a known stationary process. Section 11.1 describes simulation of autoregressive moving average (ARMA) processes with known parameters $\phi_{p,1}, \dots, \phi_{p,p}, \theta_{q,1}, \dots, \theta_{q,q}$ and σ_ϵ^2 (Equation (35b)) and harmonic processes with known parameters (Equations (35c) and (35d)). Section 11.2 discusses simulation of a stationary process for which we know a portion of its autocovariance sequence (ACVS), while Section 11.3 considers the case in which the SDF is fully specified and readily available. We next adjust procedures described in these sections to work with SDF estimates based on actual time series that we are interested in mimicking. Section 11.4 does this for nonparametric estimates (including direct SDF estimates, lag window estimates, multitaper estimates and Welch’s overlapped segment averaging [WOSA] estimates), while Section 11.5 does it for autoregressive-based parametric SDF estimates. Section 11.6 has examples of simulated series that resemble actual time series. Up to here we assume Gaussianity for the most part. Section 11.7 comments on simulation of non-Gaussian time series. We close with a summary and exercises (Sections 11.8 and 11.9).

11.1 Simulation of ARMA Processes and Harmonic Processes

In Section 2.6 we looked at examples of realizations from several discrete parameter stationary processes including white noise (see Figure 31), moving average (MA) processes (see Equation (32a) and Figure 33), autoregressive (AR) processes (Equation (33) and Figure 34) and harmonic processes (Equations (35c) and (35d) and Figure 36). Here we discuss how to generate a realization of a contiguous portion X_0, X_1, \dots, X_{N-1} from each of these processes assuming that we have access to a suitable number of independent realizations of random variables (RVs) with certain distributions. White noise, MA and AR processes are all special cases of autoregressive moving average (ARMA) processes (Equation (35b)). We tackle these special cases first, after which we turn to ARMA processes in general and then finally to harmonic processes. For convenience we take the process mean μ for the stationary processes to be zero (a realization of a process with $\mu \neq 0$ can be easily generated by taking a realization with zero mean and adding the desired nonzero μ to each of the N values constituting the realization). We also take the sampling interval Δ_t to be unity; item [3] in the Comments and Extensions (C&Es) discusses its role in the simulations.

White Noise Process

A white noise process with zero mean is a sequence of uncorrelated RVs $\{X_t : t \in \mathbb{Z}\}$, and hence $E\{X_t\} = 0$ and $\text{var}\{X_t\} = \sigma^2$ for all t , where $0 < \sigma^2 < \infty$. Since independent RVs are also uncorrelated, we can generate a realization of length N from a white noise process by simply generating N independent RVs, each of which has mean zero and variance σ^2 . Plots (a) to (f) of Figure 31 show six realizations of length $N = 100$ from white noise processes, each with zero mean and unit variance and each derived from a different selection of independent RVs (see the figure caption for details). Creation of the first four realizations requires drawing independent samples from RVs with (a) a Gaussian (normal) distribution; (b) a uniform distribution over the interval $[-\sqrt{3}, \sqrt{3}]$; (c) an exponential distribution together with a two-point distribution taking on the values ± 1 with equal probability (used to convert an exponential RV into a double exponential RV); and (d) a three-point distribution over $-5, 0$ and 5 with probabilities $0.02, 0.96$ and 0.02 . Additionally a four-point distribution over four categories with equal probability is used to create the realization in Figure 31(e) by taking a deviate from the distributions associated with the realizations in (a) to (d). In the R language the functions `rnorm`, `runif`, `rexp` and `sample` can do the random draws, but equivalent functions are available in other languages (e.g., in MATLAB the functions `normrnd`, `unifrnd` and `expnrnd` give draws from normal, uniform and exponential distributions).

Moving Average Process

Given a zero mean white noise process $\{\epsilon_t : t \in \mathbb{Z}\}$ with variance σ_ϵ^2 and given a set of $q \geq 1$ constants $\theta_{q,1}, \dots, \theta_{q,q}$ (arbitrary except for the constraint $\theta_{q,q} \neq 0$), we can construct a zero mean q th-order MA process $\{X_t : t \in \mathbb{Z}\}$ using

$$X_t = \epsilon_t - \theta_{q,1}\epsilon_{t-1} - \dots - \theta_{q,q}\epsilon_{t-q} = \sum_{j=0}^q \vartheta_{q,j}\epsilon_{t-j}, \quad (594)$$

where $\vartheta_{q,0} \stackrel{\text{def}}{=} 1$ and $\vartheta_{q,j} \stackrel{\text{def}}{=} -\theta_{q,j}$ for $j = 1, \dots, q$. Once we have a realization of $\epsilon_{-q}, \epsilon_{-q+1}, \dots, \epsilon_{N-1}$, we can readily turn it into a realization of X_0, X_1, \dots, X_{N-1} using the formula above. The summation in the formula is a convolution, which is often most efficiently computed using the discrete Fourier transform (DFT) in the following manner. Let M be any integer such that $M \geq N + q$ (DFTs are usually computed using a fast Fourier transform (FFT) algorithm, some of which work best when M is set to a power of two). Define $\vartheta_{q,j} = 0$

for $j = q + 1, q + 2, \dots, M - 1$, and define

$$\varepsilon_t = \begin{cases} \epsilon_t, & 0 \leq t \leq N - 1; \\ 0, & N \leq t \leq M - q - 1 \text{ if } M > N + q; \\ \epsilon_{t-M}, & M - q \leq t \leq M - 1. \end{cases}$$

Using the definition stated in Equation (91b), the DFTs of $\{\vartheta_{q,j}\}$ and $\{\varepsilon_t\}$ are

$$\sum_{j=0}^{M-1} \vartheta_{q,j} e^{-i2\pi k j/M} \stackrel{\text{def}}{=} \Theta_k \quad \text{and} \quad \sum_{t=0}^{M-1} \varepsilon_t e^{-i2\pi k t/M} \stackrel{\text{def}}{=} E_k, \quad k = 0, 1, \dots, M - 1.$$

The desired simulated time series is the realization of

$$X_t = \frac{1}{M} \sum_{k=0}^{M-1} \Theta_k E_k e^{i2\pi k t/M}, \quad t = 0, 1, \dots, N - 1$$

(this result follows from Equations (92a), (101d) and (101e) along with an argument similar to what is needed to solve Exercise [3.20b]). Note that we require an uncorrelated series of length $N + q$ in order to simulate a correlated series of length N . Figure 33 shows eight realizations of length $N = 128$ from MA processes of order $q = 1$ defined by $X_t = \epsilon_t - \theta_{1,1}\epsilon_{t-1}$, where $\{\epsilon_t\}$ is Gaussian white noise, and $\theta_{1,1}$ is set to 1 for four realizations (top two rows) and to -1 for the remaining four (bottom two).

Autoregressive Process

Given a zero mean white noise process $\{\epsilon_t\}$ with variance σ_p^2 and given a set of $p \geq 1$ constants $\phi_{p,1}, \dots, \phi_{p,p}$ obeying certain constraints (including $\phi_{p,p} \neq 0$ – see Section 9.2 for details), we can construct a zero mean p th-order causal (and hence stationary) AR process

$$X_t = \phi_{p,1}X_{t-1} + \phi_{p,2}X_{t-2} + \dots + \phi_{p,p}X_{t-p} + \epsilon_t = \sum_{j=1}^p \phi_{p,j}X_{t-j} + \epsilon_t. \quad (595)$$

Simulation of an AR process directly from the above equation is problematic: to simulate X_0 , we need realizations of X_{-1}, \dots, X_{-p} , but these in turn require realizations of X_{-p-1}, \dots, X_{-2p} and so forth ad infinitum. To get around this startup problem, a commonly advocated strategy is to force the realizations of X_{-1}, \dots, X_{-p} to be zero, and then use Equation (595) with realizations of $\epsilon_0, \dots, \epsilon_M, \epsilon_{M+1}, \dots, \epsilon_{N+M-1}$ for some $M > 0$ to generate realizations of $X_0, \dots, X_M, X_{M+1}, \dots, X_{N+M-1}$ (hence X_0 is set to ϵ_0 ; X_1 , to $\phi_{p,1}X_0 + \epsilon_1$, etc.). The desired simulated time series of length N is taken to be $X_M, X_{M+1}, \dots, X_{N+M-1}$, with the idea that, if M is sufficiently large, the initial zero settings should have a negligible effect on the simulated series (see the discussion on asymptotic stationarity in C&E [3] for Section 9.2). The difficulty with this burn-in procedure is that it is not usually clear what M should be set to: the larger M is, the smaller the effect of the initial zeros, but a large M can lead to an unacceptable increase in computational complexity if many realizations of an AR process are needed for a Monte Carlo experiment.

Kay (1981b) describes a procedure for simulating a Gaussian AR process that solves the startup problem by generating *stationary initial conditions* (see Section 9.4 for details on what follows). The procedure falls out from unraveling the prediction errors $\vec{\epsilon}_t(t)$ associated with the best linear predictor $\vec{X}_t(t)$ of X_t given all values of the process with time indices ranging from 0 up to $t - 1$, where t varies from 1 up to $N - 1$; for the special case $t = 0$, we define $\vec{X}_0(0)$ to be 0 (i.e., the expected value of X_0) and take $\vec{\epsilon}_0(0)$ to be X_0 . A key

property of these errors is that they are pairwise uncorrelated (see Exercise [465]). In addition $E\{\vec{\epsilon}_t(t)\} = 0$ for all t , so the prediction errors would constitute a segment of length N of zero mean Gaussian white noise were it not for the fact that their variances are not the same. Assuming $N \geq p + 1$ for convenience, the best linear predictor of X_t for $t = p, \dots, N - 1$ is given by the right-hand side of Equation (595) with ϵ_t replaced by its expected value of zero:

$$\vec{X}_t(t) \stackrel{\text{def}}{=} \sum_{j=1}^p \phi_{p,j} X_{t-j}.$$

The associated prediction error is

$$\vec{\epsilon}_t(t) \stackrel{\text{def}}{=} X_t - \vec{X}_t(t) = \epsilon_t,$$

and hence $\text{var}\{\vec{\epsilon}_t(t)\} = \sigma_p^2$. For $t = 0, \dots, p - 1$, we can't appeal to Equation (595) directly to forecast X_t since we have fewer than p values prior to X_t . The best linear predictors are given by $\vec{X}_0(0) = 0$ and, if $p \geq 2$, by

$$\vec{X}_t(t) \stackrel{\text{def}}{=} \sum_{j=1}^t \phi_{t,j} X_{t-j}, \quad t = 1, \dots, p - 1,$$

(see Equation (452)). Let $\sigma_0^2, \sigma_1^2, \dots, \sigma_{p-1}^2$ denote the variances of the associated prediction errors, which, starting with σ_p^2 and using $\phi_{p,p}$ from Equation (595), can be computed recursively via

$$\sigma_{t-1}^2 = \frac{\sigma_t^2}{1 - \phi_{t,t}^2}, \quad t = p, p - 1, \dots, 1, \quad (596a)$$

(see Equation (460b)). Given a realization of a portion Z_0, Z_1, \dots, Z_{N-1} of a Gaussian white noise process with zero mean and unit variance, we can form a realization of the prediction errors $\vec{\epsilon}_t(t)$, $t = 0, 1, \dots, N - 1$, by equating them to realizations of $\sigma_0 Z_0, \dots, \sigma_{p-1} Z_{p-1}, \sigma_p Z_p, \dots, \sigma_p Z_{N-1}$. Now note that the system of N equations

$$\begin{aligned} \vec{\epsilon}_0(0) &= X_0 \\ \vec{\epsilon}_1(1) &= X_1 - \phi_{1,1} X_0 \quad (\text{not needed if } p = 1) \\ \vec{\epsilon}_2(2) &= X_2 - \sum_{j=1}^2 \phi_{2,j} X_{2-j} \quad (\text{not needed if } p = 1 \text{ or } 2) \\ &\vdots \\ \vec{\epsilon}_{p-1}(p-1) &= X_{p-1} - \sum_{j=1}^{p-1} \phi_{p-1,j} X_{p-1-j} \quad (\text{not needed if } p = 1, 2 \text{ or } 3) \\ \vec{\epsilon}_t(t) &= \epsilon_t = X_t - \sum_{j=1}^p \phi_{p,j} X_{t-j}, \quad t = p, \dots, N - 1 \end{aligned} \quad (596b)$$

can be reexpressed – upon replacing the prediction errors with distributionally equivalent RVs – as

$$\begin{aligned} X_0 &= \sigma_0 Z_0 \\ X_1 &= \phi_{1,1} X_0 + \sigma_1 Z_1 \quad (\text{not needed if } p = 1) \end{aligned}$$

$$\begin{aligned}
X_2 &= \sum_{j=1}^2 \phi_{2,j} X_{2-j} + \sigma_2 Z_2 \quad (\text{not needed if } p = 1 \text{ or } 2) \\
&\vdots \\
X_{p-1} &= \sum_{j=1}^{p-1} \phi_{p-1,j} X_{p-1-j} + \sigma_{p-1} Z_{p-1} \quad (\text{not needed if } p = 1, 2 \text{ or } 3) \\
X_t &= \sum_{j=1}^p \phi_{p,j} X_{t-j} + \sigma_p Z_t, \quad t = p, \dots, N-1,
\end{aligned} \tag{597a}$$

which gives us a recursive way of transforming RVs that have the same multivariate distribution as the uncorrelated prediction errors into a correlated AR series.

To implement this scheme, we assume knowledge of the $\phi_{p,j}$ coefficients, but we also need the $\phi_{t,j}$ coefficients for t less than p when $p \geq 2$. Thus, when $p = 1$, all that is required is $\phi_{1,1}$, i.e., the AR(1) coefficient. When $p \geq 2$, starting with the AR coefficients $\phi_{p,j}$ and with t set to p , we can use

$$\phi_{t-1,j} = \frac{\phi_{t,j} + \phi_{t,t}\phi_{t,t-j}}{1 - \phi_{t,t}^2}, \quad 1 \leq j \leq t-1, \tag{597b}$$

to obtain the required coefficients for order $p-1$ (see Equation (460a)). With these coefficients, we can use the above again if need be to obtain the coefficients for order $p-2$ and so forth, ending the recursions with the computation of $\phi_{1,1}$.

As an example, let us consider the second-order Gaussian AR process of Equation (34), namely,

$$X_t = \frac{3}{4}X_{t-1} - \frac{1}{2}X_{t-2} + \epsilon_t \quad \text{with } \sigma_2^2 = 1.$$

▷ **Exercise [597]** Show that, given a realization of a portion Z_0, Z_1, \dots, Z_{N-1} of Gaussian white noise with zero mean and unit variance, we can generate a realization of length N of the AR process above via

$$\begin{aligned}
X_0 &= \frac{4}{3}Z_0 \\
X_1 &= \frac{1}{2}X_0 + \frac{2}{\sqrt{3}}Z_1 \\
X_t &= \frac{3}{4}X_{t-1} - \frac{1}{2}X_{t-2} + Z_t, \quad t = 2, \dots, N-1.
\end{aligned} \tag{597c}$$

For a second example, see Exercise [11.1], which considers the fourth-order Gaussian AR process of Equation (35a).

Figure 34 shows eight realizations of AR processes, each of length $N = 1024$. The top two rows show four realizations of the AR(2) process of Equation (34), and the bottom two, of the AR(4) process of Equation (35a).

Autoregressive Moving Average Process

A zero mean ARMA process $\{X_t\}$ of order $p \geq 1$ and $q \geq 1$ combines the AR process of Equation (595) with the MA process of Equation (594) by using the latter as a replacement for the white noise process driving the former:

$$X_t = \sum_{j=1}^p \phi_{p,j} X_{t-j} + \sum_{j=0}^q \vartheta_{q,j} \epsilon_{t-j}. \tag{597c}$$

▷ **Exercise [598]** Show that an ARMA process can be decomposed as follows (Kay, 1981b):

$$X_t = \sum_{j=0}^q \vartheta_{q,j} Y_{t-j}, \quad \text{where } Y_t \stackrel{\text{def}}{=} \sum_{k=1}^p \phi_{p,k} Y_{t-k} + \epsilon_t. \quad \triangleleft$$

We can thus simulate a segment of length N from a Gaussian ARMA process by first simulating a segment of length $N + q$ from the AR process $\{Y_t\}$ – to be treated as a realization of $Y_{-q}, Y_{-q+1}, \dots, Y_{N-1}$ – and then following the recipe for simulating an MA process, but with the AR realization replacing the realization of $\epsilon_{-q}, \epsilon_{-q+1}, \dots, \epsilon_{N-1}$.

Harmonic Process

A zero mean harmonic process $\{X_t\}$ can be written as

$$X_t = \sum_{l=1}^L A_l \cos(2\pi f_l t) + B_l \sin(2\pi f_l t), \quad t \in \mathbb{Z}, \quad (598a)$$

where f_l are real-valued positive constants, L is a positive integer, and A_l and B_l are uncorrelated real-valued RVs with zero means such that $\text{var}\{A_l\} = \text{var}\{B_l\}$ (see Equation (35c)). To create a realization of X_0, X_1, \dots, X_{N-1} , we need only generate uncorrelated realizations of the $2L$ RVs A_l and B_l , and then evaluate the right-hand side of the above for $t = 0, 1, \dots, N - 1$. If each of the $2L$ RVs obeys a Gaussian distribution, then the process $\{X_t\}$ is also Gaussian with mean zero and variance $\sum_{l=1}^L \text{var}\{A_l\}$ (see Equation (36b)). The left-hand column of Figure 36 shows three realizations of a Gaussian harmonic process with $L = 1$ and $f_1 = 1/20$ for which the realizations of A_1 and B_1 come from a standard Gaussian distribution. Since we only need that the RVs A_l and B_l be uncorrelated, and since independence implies uncorrelatedness, we could simulate a non-Gaussian harmonic process by, e.g., generating realizations of $2L$ RVs, each with a possibly different distribution with the only restriction being that $\text{var}\{A_l\}$ and $\text{var}\{B_l\}$ must be equal for a given l .

We can reexpress the harmonic process of Equation (598a) as

$$X_t = \sum_{l=1}^L D_l \cos(2\pi f_l t + \phi_l), \quad t \in \mathbb{Z}, \quad (598b)$$

where $D_l^2 = A_l^2 + B_l^2$, $\tan(\phi_l) = -B_l/A_l$, $A_l = D_l \cos(\phi_l)$ and $B_l = -D_l \sin(\phi_l)$ (see Equation (35d)). As noted following that equation, if the A_l and B_l RVs are all Gaussian, then the D_l^2 RVs are independent with exponential PDFs, but these RVs need not be identically distributed. The mean completely characterizes the exponential distribution, and the mean for the l th RV D_l^2 is $2 \text{var}\{A_l\}$, which need not be the same for all l . The nonnegative random amplitude D_l is said to obey a Rayleigh distribution, so the D_l RVs are independent and Rayleigh distributed, but they need not be identically distributed. On the other hand, the phases ϕ_l are independent and identically distributed RVs. Their common distribution is uniform over the interval $(-\pi, \pi]$, and the ϕ_l RVs are independent of the D_l RVs.

To simulate a Gaussian harmonic process that obeys Equation (598b) rather than Equation (598a), we need to generate L independent realizations of RVs with exponential PDFs and with means given by $2 \text{var}\{A_l\}$. The square roots of these realizations are realizations of the random amplitudes D_l . Independent of these L realizations, we also need to generate realizations of L independent and identically distributed (IID) RVs that are uniformly distributed over $(-\pi, \pi]$ – these realizations serve as realizations of the ϕ_l RVs. To generate harmonic processes that need not obey a Gaussian distribution for all t , it is sufficient to generate the L

random phases as before, but the L independent random amplitudes can be any set of RVs that yield positive values with probability one. In particular, the amplitudes can be fixed, i.e., have degenerate distributions. Three realizations from a harmonic process with fixed amplitudes but random phases are shown in the right-hand column of Figure 36 for the case of $L = 1$, $f_1 = 1/20$ and $D_1 = \sqrt{2}$. The ACVS for this non-Gaussian harmonic process is the same as that of the Gaussian-distributed harmonic process used to generate the realizations in the left-hand column.

Comments and Extensions to Section 11.1

[1] In defining an AR process $\{X_t\}$, we have taken the variance σ_p^2 of the associated white noise process to be one of the $p + 1$ parameters. It is sometimes convenient to use the process variance $\text{var}\{X_t\}$ instead. Noting that $\sigma_0^2 = \text{var}\{\vec{\epsilon}_0(0)\} = \text{var}\{X_0 - \vec{X}_0(0)\} = \text{var}\{X_t\}$ since $\vec{X}_0(0) = 0$, we can readily obtain the remaining prediction error variances required for the simulation scheme by reexpressing Equation (596a) as

$$\sigma_t^2 = \sigma_{t-1}^2 (1 - \phi_{t,t}^2), \quad t = 1, 2, \dots, p.$$

[2] The approach presented above for generating realizations of Gaussian ARMA processes with stationary initial conditions follows Kay (1981b). McLeod and Hipel (1978) describe another method for doing so that has some interesting contrasts with Kay's approach. For an MA process, their method is the same as what we have presented. For a p th-order Gaussian AR process, the first step in their method is to form the covariance matrix \mathbf{I}_p for X_0, \dots, X_{p-1} ; i.e., its (j, k) th element is s_{j-k} , where $\{s_\tau\}$ is the ACVS for the AR process – cf. Equation (450c). All $p \times p$ entries of \mathbf{I}_p are known once we know s_0, \dots, s_{p-1} . Given the $p + 1$ parameters $\phi_{p,1}, \dots, \phi_{p,p}$ and σ_p^2 for the process, part of the burden of Exercise [9.11] is to show how to obtain these ACVS values. The second step calls upon a general purpose routine to construct the Cholesky decomposition of \mathbf{I}_p (see, e.g., Ralston and Rabinowitz, 1978, or Golub and Van Loan, 2013). This yields a lower triangular matrix \mathbf{C}_p such that $\mathbf{C}_p \mathbf{C}_p^T = \mathbf{I}_p$. Let \mathbf{Z}_p be a p -dimensional vector containing p IID standard Gaussian RVs. A standard result in the theory of Gaussian vectors says that $\mathbf{C}_p \mathbf{Z}_p$ has the same distribution as X_0, \dots, X_{p-1} (see, e.g., section A.3 of Brockwell and Davis, 2016). Hence a realization of $\mathbf{C}_p \mathbf{Z}_p$ gives us a realization of X_0, \dots, X_{p-1} . Given these stationary initial conditions, we can in turn generate realizations of X_p, \dots, X_{N-1} by generating realizations of $N - p$ additional IID standard Gaussian RVs, multiplying these by σ_p so that they can be regarded as realizations of $\epsilon_p, \dots, \epsilon_{N-1}$ and then making use of the defining equation for an AR process (Equation (595)).

For the AR case, what is the difference between the methods of Kay (1981b) and of McLeod and Hipel (1978)? The two methods simulate X_p, \dots, X_{N-1} in exactly the same way once they have generated the stationary initial conditions for $\mathbf{X}_p \stackrel{\text{def}}{=} [X_0, \dots, X_{p-1}]^T$. McLeod and Hipel generate the conditions via $\mathbf{X}_p \stackrel{\text{d}}{=} \mathbf{C}_p \mathbf{Z}_p$, which uses the Cholesky decomposition $\mathbf{I}_p = \mathbf{C}_p \mathbf{C}_p^T$ (recall that $\stackrel{\text{d}}{=}$ means “equal in distribution”). In contrast, Kay uses the modified Cholesky decomposition $\mathbf{I}_p = \mathbf{L}_p \mathbf{D}_p \mathbf{L}_p^T$, where \mathbf{L}_p is a $p \times p$ lower triangular matrix whose diagonal elements are all equal to 1, and \mathbf{D}_p is a $p \times p$ diagonal matrix whose diagonal elements are all positive (see the discussion surrounding Equation (464c)). Let $\mathbf{D}_p^{1/2}$ be a diagonal matrix whose diagonal elements are the square roots of those in \mathbf{D}_p , and hence $\mathbf{D}_p^{1/2} \mathbf{D}_p^{1/2} = \mathbf{D}_p$. It follows that the Cholesky and modified Cholesky decompositions are related by $\mathbf{C}_p = \mathbf{L}_p \mathbf{D}_p^{1/2}$. Let $\vec{\epsilon}_p \stackrel{\text{def}}{=} [\vec{\epsilon}_0(0), \dots, \vec{\epsilon}_{p-1}(p-1)]^T$. Kay's method for generating stationary initial conditions is based on $\mathbf{L}_p^{-1} \mathbf{X}_p = \vec{\epsilon}_p$ (see Equation (465a)). Since $\vec{\epsilon}_p \stackrel{\text{d}}{=} \mathbf{D}_p^{1/2} \mathbf{Z}_p$, we have $\mathbf{L}_p^{-1} \mathbf{X}_p \stackrel{\text{d}}{=} \mathbf{D}_p^{1/2} \mathbf{Z}_p$ and hence $\mathbf{X}_p \stackrel{\text{d}}{=} \mathbf{L}_p \mathbf{D}_p^{1/2} \mathbf{Z}_p = \mathbf{C}_p \mathbf{Z}_p$, which is the same as the basis for the McLeod and Hipel method. The two methods thus produce equivalent stationary initial conditions. The theoretical advantage of Kay's method is that the elements of \mathbf{L}_p^{-1} and $\mathbf{D}_p^{1/2}$ are easily obtained

from the $p + 1$ AR parameters via Equations (597b) and (596a) because

$$\mathbf{L}_p^{-1} = \begin{bmatrix} 1 & & & \\ -\phi_{1,1} & 1 & & \\ -\phi_{2,2} & -\phi_{2,1} & 1 & \\ \vdots & \vdots & \ddots & \\ -\phi_{p-1,p-1} & \dots & -\phi_{p-1,1} & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{D}_p^{1/2} = \begin{bmatrix} \sigma_0 & & & \\ & \sigma_1 & & \\ & & \sigma_2 & \\ & & & \ddots \\ & & & & \sigma_{p-1} \end{bmatrix}$$

(see Equation (465b) and the description following it). In addition, there is no need to get \mathbf{L}_p explicitly by inverting \mathbf{L}_p^{-1} since we can obtain the elements of the stationary initial conditions $\mathbf{L}_p \mathbf{D}_p^{1/2} \mathbf{Z}_p$ by the unravelling procedure of Equation (597a).

Turning now to Gaussian ARMA processes, McLeod and Hipel (1978) generate stationary initial conditions based upon the joint distribution of the $p + q$ RVs $X_0, \dots, X_{p-1}, \epsilon_{p-q}, \dots, \epsilon_{p-1}$ (once realizations of these RVs are provided, it is an easy matter to generate realizations of X_p, \dots, X_{N-1} via Equation (597c) by generating realizations of $\epsilon_p, \dots, \epsilon_{N-1}$). The covariance matrix for these $p + q$ RVs is given by

$$\boldsymbol{\Sigma}_{p+q} = \begin{bmatrix} \mathbf{I}_p & \sigma_\epsilon^2 \boldsymbol{\Psi}_{p,q} \\ \sigma_\epsilon^2 \boldsymbol{\Psi}_{p,q}^T & \sigma_\epsilon^2 \mathbf{I}_q \end{bmatrix}, \quad (600a)$$

where the (j, k) th entry of the $p \times p$ matrix \mathbf{I}_p is s_{j-k} , and s_{j-k} is the ACVS of the ARMA process; the (t, k) th entry of the $p \times q$ matrix $\sigma_\epsilon^2 \boldsymbol{\Psi}_{p,q}$ is $\text{cov}\{X_t, \epsilon_{p-q+k}\}$, where $t = 0, \dots, p-1$ and $k = 0, \dots, q-1$; and $\sigma_\epsilon^2 \mathbf{I}_q$ is the covariance matrix for $\epsilon_{p-q}, \dots, \epsilon_{p-1}$, where \mathbf{I}_q is the $q \times q$ identity matrix. Section 3.2.1 of Brockwell and Davis (2016) gives three methods for determining the ACVS for an ARMA process from its $p + q + 1$ parameters. Here is a fourth based upon the decomposition in Exercise [598] in conjunction with the method for determining the ACVS for an AR process given by Exercise [9.11]. Let $\{s_{Y,\tau}\}$ denote the ACVS for AR process $\{Y_t\}$ defined in Exercise [598]. An application of Exercise [2.1e] shows that the ACVS $\{s_\tau\}$ for the ARMA process $\{X_t\}$ is given by

$$s_\tau = \sum_{j=0}^q \sum_{k=0}^q \vartheta_{q,j} \vartheta_{q,k} s_{Y,\tau-j+k}. \quad (600b)$$

To form the elements of \mathbf{I}_p , we need to evaluate $s_{Y,\tau}$ at lags $\tau = 0, \dots, p + q - 1$ in order to get s_τ at the required lags $\tau = 0, \dots, p - 1$. We next need to determine $\sigma_\epsilon^2 \boldsymbol{\Psi}_{p,q}$. If we assume that the ARMA process is causal, there exists a unique set of weights $\{\psi_j\}$ such that

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \quad (600c)$$

where $\psi_0 \stackrel{\text{def}}{=} 1$ and $\psi_j \stackrel{\text{def}}{=} 0$ for $j < 0$ (as noted in section 3.1 of Brockwell and Davis, 2016, an ARMA process is causal if the AR coefficients satisfy the same condition that we need for an AR process to be causal; i.e., the roots of the polynomial $1 - \sum_{j=1}^p \phi_{p,j} z^{-j}$ must all lie inside the unit circle, and no constraints are needed on the MA coefficients). Hence

$$\text{cov}\{X_t, \epsilon_{p-q+k}\} = \sum_{j=-\infty}^{\infty} \psi_j E\{\epsilon_{t-j} \epsilon_{p-q+k}\} = \sigma_\epsilon^2 \psi_{t-p+q-k};$$

however, since $t = 0, \dots, p-1$ and $k = 0, \dots, q-1$, we actually only need to compute ψ_j for $j = 0, \dots, q-1$ to fill in the nonzero entries of $\boldsymbol{\Psi}_{p,q}$. A recursive scheme for doing so is

$$\psi_j = \sum_{k=1}^p \phi_{p,k} \psi_{j-k} + \vartheta_{q,j}, \quad j = 0, \dots, q-1 \quad (600d)$$

(Brockwell and Davis, 2016, equation (3.1.7)). With Σ_{p+q} now completely determined, we form its Cholesky decomposition $\Sigma_{p+q} = \mathbf{C}_{p+q} \mathbf{C}_{p+q}^T$ and can then obtain a realization of $X_0, \dots, X_{p-1}, \epsilon_{p-q}, \dots, \epsilon_{p-1}$ via $\mathbf{C}_{p+q} \mathbf{Z}_{p+q}$, where \mathbf{Z}_{p+q} is a vector containing $p+q$ standard Gaussian RVs. Exercise [11.2] invites the reader to work out the details of this method for the case $p=1$ and $q=1$ and to compare it to Kay's method.

[3] As noted in Section 2.2, the t in X_t is a unitless index and not an actual time. We can take the time associated with X_t to be $t_0 + t \Delta_t$, where t_0 is the actual time at which X_0 is observed, and $\Delta_t > 0$ is the sampling interval between adjacent values of the time series, e.g., X_t and X_{t+1} . The procedures we have described for generating ARMA processes – and special cases thereof – yield realizations of X_0, X_1, \dots, X_{N-1} that we are free to associate with times dictated by arbitrary choices for t_0 and $\Delta_t > 0$. The same is true for harmonic processes if we are willing to assume that the positive frequencies f_t in Equations (598a) and (598b) are unitless. If we want them to have units, e.g., Hz, then we need to replace t in the summations in these equations with $t \Delta_t$, where $\Delta_t = 1$ sec for this example. With this replacement, we can associate X_0, X_1, \dots, X_{N-1} with times $t_0, t_0 + \Delta_t, \dots, t_0 + (N-1) \Delta_t$ for any choice of t_0 so desired.

11 Simulation of Processes with a Known Autocovariance Sequence

Here we consider how to generate a realization of a portion X_0, X_1, \dots, X_{N-1} of a Gaussian stationary process given that we know part of its ACVS $\{s_\tau\}$ (without loss of generality, we assume that the process mean for $\{X_t\}$ is zero and that its sampling interval Δ_t is unity). Under the Gaussian assumption, a particularly attractive approach is *circulant embedding*, which is also known as the Davies–Harte method since it was introduced into the statistical literature in an appendix to Davies and Harte (1987); however, the method predates this appendix, with Woods (1972) being an early reference outside the context of time series. Subject to certain conditions (stated in the next paragraph), the method is exact in the sense that it produces simulated time series whose statistical properties are *exactly* the same as those of the desired X_0, X_1, \dots, X_{N-1} . The method is not entirely general because there are certain stationary processes and certain sample sizes N for which the required conditions do not hold. In what follows, we present a formulation of circulant embedding due to Dietrich and Newsam (1997), which is an appealing alternative to that of Davies and Harte (1987).

Assume that s_0, s_1, \dots, s_N are known, and compute the weights

$$S_k \stackrel{\text{def}}{=} \sum_{\tau=0}^N s_\tau e^{-i2\pi \tilde{f}_k \tau} + \sum_{\tau=N+1}^{2N-1} s_{2N-\tau} e^{-i2\pi \tilde{f}_k \tau}, \quad k = 0, 1, \dots, 2N-1, \quad (601a)$$

where $\tilde{f}_k = k/(2N)$ defines a grid of frequencies twice as fine as the Fourier frequencies associated with a time series of length N . Equation (91b) tells us that $\{S_k\}$ is the DFT of

$$s_0, s_1, \dots, s_{N-2}, s_{N-1}, s_N, s_{N-1}, s_{N-2}, \dots, s_1. \quad (601b)$$

This “circularized” sequence of length $2N$ has “embedded” within it the $N+1$ values of the ACVS that are assumed to be known.

▷ **Exercise [601]** Show that the weights S_k are real-valued. ◁

For circulant embedding to work, the weights S_k *must* be nonnegative. This condition is known to hold for certain stationary processes, but not all (see, e.g., Gneiting, 2000, Craigmille, 2003, and references therein). Assuming that $S_k \geq 0$ for all k , let $Z_0, Z_1, \dots, Z_{4N-1}$ be a set of $4N$ independent standard Gaussian RVs, and define

$$\mathcal{Y}_k = \left(\frac{S_k}{2N} \right)^{1/2} (Z_{2k} + iZ_{2k+1}), \quad 0 \leq k \leq 2N-1. \quad (601c)$$

Let $\{Y_t\}$ be the DFT of $\{\mathcal{Y}_k\}$:

$$Y_t = \sum_{k=0}^{2N-1} \mathcal{Y}_k e^{-i2\pi \tilde{f}_k t}. \quad (602)$$

The process $\{Y_t\}$ is complex-valued and well defined for all $t \in \mathbb{Z}$ (it is a periodic sequence with a period of $2N$). Let $Y_{\Re,t} \stackrel{\text{def}}{=} \Re\{Y_t\}$ and $Y_{\Im,t} \stackrel{\text{def}}{=} \Im\{Y_t\}$, where, as usual, $\Re\{z\}$ and $\Im\{z\}$ are the real and imaginary components of the complex-valued variable z .

- ▷ **Exercise [602]** Show that the real-valued processes $\{Y_{\Re,t}\}$ and $\{Y_{\Im,t}\}$ are zero mean Gaussian stationary processes, both of which have ACVSs that are in agreement with that of $\{X_t\}$ at lags $\tau = 0, 1, \dots, N$, but not necessarily at lags $\tau > N$; moreover, the RVs in $\{Y_{\Re,t}\}$ are independent of the RVs in $\{Y_{\Im,t}\}$. ◁

We can thus regard realizations of $Y_{\Re,0}, Y_{\Re,1}, \dots, Y_{\Re,N-1}$ and $Y_{\Im,0}, Y_{\Im,1}, \dots, Y_{\Im,N-1}$ as two independent realizations of X_0, X_1, \dots, X_{N-1} (actually this statement holds if we replace $N-1$ with N – see C&E [1] later in this section for details).

In summary, given knowledge of a portion $s_\tau, \tau = 0, 1, \dots, N$, of the ACVS for a Gaussian stationary process $\{X_t\}$, the circulant embedding method consists of

- [1] computing the DFT $\{S_k\}$ of the circularized sequence given in Equation (601b);
- [2] checking that $S_k \geq 0$ for $k = 0, 1, \dots, 2N-1$;
- [3] if the S_k weights are nonnegative, using $4N$ deviates from a Gaussian white noise process $\{Z_t\}$ with zero mean and unit variance to form \mathcal{Y}_k as per Equation (601c); and, finally,
- [4] taking the DFT of $\{\mathcal{Y}_k\}$ to obtain $Y_t = Y_{\Re,t} + iY_{\Im,t}$ – we can consider realizations of the real-valued sequences $Y_{\Re,t}$ and $Y_{\Im,t}, t = 0, 1, \dots, N-1$, to be two independent realizations of $X_t, t = 0, 1, \dots, N-1$.

We can compute the DFTs needed in steps [1] and [4] efficiently if $2N$ is amenable to an FFT algorithm. We can obtain as many additional pairs of realizations of X_t as desired by repeating steps [3] and [4] over and over again, with a different random sample of Gaussian white noise used at each iteration of step [3].

As an example, let us generate realizations of length $N = 1024$ from the AR(2) process of Equation (34) using circulant embedding (we do this for illustrative purposes only – Exercise [597] gives the preferred method). To compute the required weights S_k using Equation (601a), we need the ACVS $\{s_\tau\}$ for the AR(2) process at lags $\tau = 0, 1, \dots, 1024$, which we can obtain using Equation (508a). Using these ACVS values, we form the circularized sequence of Equation (601b) and take its DFT, which yields the desired weights $S_k, k = 0, 1, \dots, 2047$. These weights are all positive (as required by the circulant embedding method) and are shown in Figure 603(a) (the smallest weight is $S_{1024} \doteq 0.19753$). Using these weights, we next create the complex values $\mathcal{Y}_k, k = 0, 1, \dots, 2047$, as per Equation (601c), for which we need 4096 independent realizations from a standard Gaussian distribution. Finally we take the DFT of $\{\mathcal{Y}_k\}$ to obtain the complex-valued sequence $\{Y_t\}$ of Equation (602), the real and imaginary parts of which are shown by the curves in Figures 603(b) and (c). The first 1024 deviates in both parts are indicated by the darker portions of the curves – these are two realizations of the AR(2) process, and these realizations are independent of each other. The appearances of both realizations are qualitatively similar to the four realizations of this process shown in the top four plots of Figure 34, which were generated using the preferred procedure stated in Exercise [597]. (The lighter portions of the curves in plots (b) and (c) have a similar appearance to the darker portions – C&E [2] for this section discusses why this is to be expected; see also Exercise [11.5], which uses this AR(2) example as a follow-on to Exercise [602]).

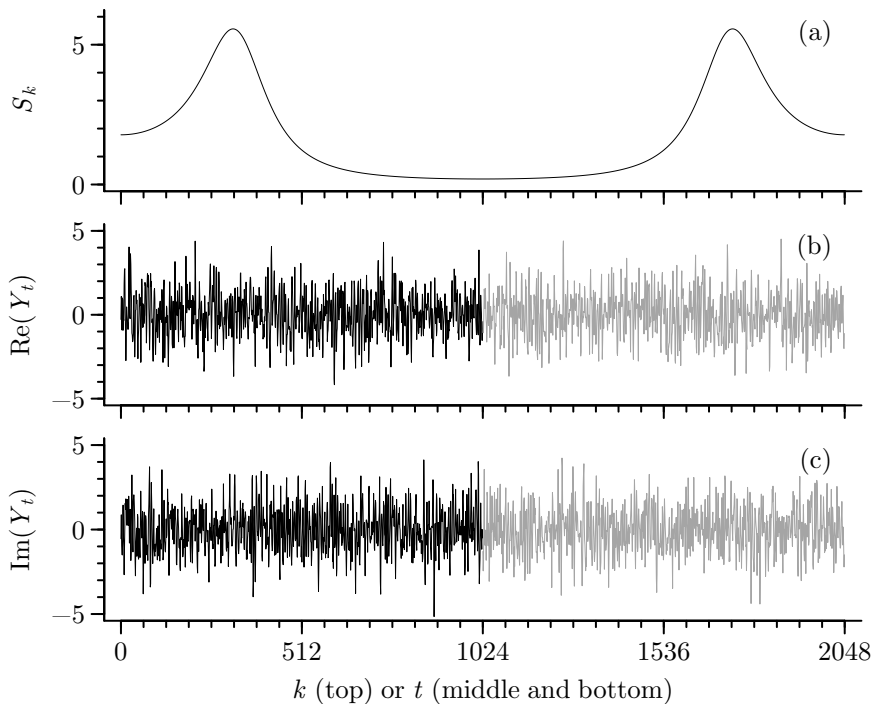


Figure 603 Creation via circulant embedding of realizations of length $N = 1024$ from the Gaussian AR(2) process of Equation (34). Plot (a) shows the weights S_k , $k = 0, 1, \dots, 2047$, used to create the complex-valued series Y_t , $t = 0, 1, \dots, 2047$, while (b) and (c) show, respectively, the real and imaginary parts of $\{Y_t\}$. The first 1024 values of both (the darker portions of the plots) constitute two independent realizations of the AR(2) process (cf. the top four plots of Figure 34, which were generated via the method detailed in Exercise [597]).

Comments and Extensions to Section 11.2

[1] Our goal in this section was to formulate an exact method for simulating a time series of length N , but the version of circulant embedding we presented yields simulated series of length $N + 1$ (one longer than desired). It is easy to achieve our goal by just ignoring $Y_{\Re, N}$ and $Y_{\Im, N}$, but, to avoid tossing anything away, why not just reformulate the method with N replaced throughout by $N - 1$ (in particular, the middle part of Equation (601b) would become $s_{N-3}, s_{N-2}, s_{N-1}, s_{N-2}, s_{N-3}$)? An appealing aspect of circulant embedding is that it can be implemented efficiently using an FFT algorithm. Because simple versions of this algorithm are designed to work with series whose lengths are powers of two, sample sizes of time series are often set such that $N = 2^J$ for some positive integer J . Our formulation of circulant embedding requires taking the DFT of a series of length $2N$, which is a power of two if N is a power of two. If we were to reformulate the method by replacing N with $N - 1$, we would need to take the DFT of a series of length $2N - 2$, which is in general not a power of two if N is such. The formulation we have presented is thus motivated by the desire to make the circulant embedding method computationally efficient when using a simple FFT algorithm with a desired sample size N that is a power of two (see C&E [1] for Section 6.3 for another example of the interplay between N and FFT algorithms).

One consequence of our formulation is that we need not insist on s_N being part of the circularized sequence of Equation (601b) – we are free to replace it with some other value, but keeping in mind the goal that the weights arising from the DFT of the modified circularized sequence should be nonnegative (see Exercise [11.4c] for an example).

[2] The solution to Exercise [602] indicates that both $\{Y_{\Re, t} : t \in \mathbb{Z}\}$ and $\{Y_{\Im, t} : t \in \mathbb{Z}\}$ are special cases of the harmonic process of Equation (35c), each being composed of sinusoids with frequencies $\pi k/N$, $k = 0, 1, \dots, 2N - 1$; moreover, these frequencies are such that both processes are periodic

with a period of $2N$ (an easy extension to the exercise). Knowledge of realizations of these processes over $0 \leq t \leq 2N - 1$ is thus equivalent to knowledge over $t \in \mathbb{Z}$. Turning to the example displayed in Figure 603 for which $N = 1024$, the realizations of $Y_{\mathcal{R},t}$ and $Y_{\mathcal{S},t}$ shown for $0 \leq t \leq 2047$ in plots (b) and (c) are also what the realizations look like over, e.g., $2048 \leq t \leq 4095$.

Since $\{Y_{\mathcal{R},t}\}$ and $\{Y_{\mathcal{S},t}\}$ are periodic, the series $Y_{\mathcal{R},0}, Y_{\mathcal{R},1}, \dots, Y_{\mathcal{R},N-1}$ and $Y_{\mathcal{S},0}, Y_{\mathcal{S},1}, \dots, Y_{\mathcal{S},N-1}$ are not the only ones we can regard as two independent realizations of X_0, X_1, \dots, X_{N-1} . In fact the series $Y_{\mathcal{R},\nu_{\mathcal{R}}}, Y_{\mathcal{R},\nu_{\mathcal{R}}+1}, \dots, Y_{\mathcal{R},\nu_{\mathcal{R}}+N-1}$ and $Y_{\mathcal{S},\nu_{\mathcal{S}}}, Y_{\mathcal{S},\nu_{\mathcal{S}}+1}, \dots, Y_{\mathcal{S},\nu_{\mathcal{S}}+N-1}$ are also such for *any* choice of $\nu_{\mathcal{R}}$ and $\nu_{\mathcal{S}}$ (in particular these integers need *not* be the same). Looking again at Figure 603, we can take the two independent realizations to be (i) the lighter series in the two plots; (ii) the lighter series in (b) in combination with the darker series in (c); (iii) the portion of the series in (b) indexed by $t = 512, 513, \dots, 1535$ together with the darker series in (c); and so forth; however, given that we cannot in general extract more than two independent realizations, there is nothing to be gained by not sticking with just $\nu_{\mathcal{R}} = \nu_{\mathcal{S}} = 0$.

[3] There are some stationary processes $\{X_t\}$ in combination with certain sample sizes N for which the weights S_k generated via Equation (601a) are not all nonnegative, hence rendering the circulant embedding method formally inapplicable (in particular, the square root of S_k in Equation (601c) poses a problem). In these cases, an alternative method for generating exact simulations makes use of the Cholesky decomposition of the $N \times N$ covariance matrix $\mathbf{\Gamma}_N$ for $\mathbf{X}_N \stackrel{\text{def}}{=} [X_0, \dots, X_{N-1}]^T$. Under the assumption that s_0, s_1, \dots, s_{N-1} are known, the matrix $\mathbf{\Gamma}_N$ is also known since its (j, k) th element is s_{j-k} , and $0 \leq |j - k| \leq N - 1$. As noted in C&E [2] for Section 11.1, the Cholesky decomposition takes the form $\mathbf{\Gamma}_N = \mathbf{C}_N \mathbf{C}_N^T$, where \mathbf{C}_N is a lower triangular matrix whose elements can be obtained efficiently via the Levinson–Durbin recursions. Simulation of \mathbf{X}_N is accomplished by multiplying \mathbf{C}_N with an N -dimensional vector \mathbf{Z}_N containing N IID standard Gaussian RVs; i.e., $\mathbf{X}_N \stackrel{\text{d}}{=} \mathbf{C}_N \mathbf{Z}_N$.

In contrast to the circulant embedding method, there are no theoretical restrictions on the Cholesky decomposition method since \mathbf{C}_N is well defined for all covariance matrices $\mathbf{\Gamma}_N$. There are, however, two practical problems that restrict its applicability. First, even with the help of the computationally efficient Levinson–Durbin recursions, there are certain processes for which computing \mathbf{C}_N is challenging. One example is band-limited white noise, whose ACVS is given by $s_{\tau}^{(\text{BL})} = \sin(2\pi W\tau)/(\pi\tau)$, where $0 < W < 1/2$ (this is Equation (379b) with $\Delta_t = 1$). We invite readers to verify that use of the Levinson–Durbin recursions to compute \mathbf{C}_N for the case $W = 1/8$ and $N = 1024$ leads to substantial numerical instabilities (in particular, we found that, while all 1024 diagonal elements of \mathbf{C}_N must in theory be nonnegative, 398 of them were incorrectly computed to be negative). Second, even if \mathbf{C}_N can be computed accurately, generating realizations via $\mathbf{C}_N \mathbf{Z}_N$ has a computational complexity of $O(N^2)$, which can render this method painfully slow even for moderate sample sizes, say $N = 1024$; by contrast, if an FFT algorithm is used, the computational complexity of the circulant embedding method is $O(N \log_2(N))$.

Since the circulant embedding method is computationally attractive, it is of interest to consider two modifications that allow its use when $S_k < 0$ for some k . First, we can generate series of longer length, say $N' > N$. If the weights for the length N' series prove to be nonnegative, we can get a series of the desired length by just extracting its first N values. Exercise [11.6a] gives an example where this works. Second, we can define a new set of weights, say \tilde{S}_k , by replacing any negative weights with zeros (i.e., $\tilde{S}_k = (S_k)_+ \stackrel{\text{def}}{=} \max\{S_k, 0\}$) and proceed as usual, but, while the resulting method is no longer exact, it is possible to measure how inaccurate it is. Exercise [11.6b] gives an example of this approach.

11.3 Simulation of Processes with a Known Spectral Density Function

In the previous section we considered simulation of a portion X_0, X_1, \dots, X_{N-1} of a zero mean Gaussian stationary process $\{X_t\}$ under the assumption that we know part of its ACVS $\{s_{\tau}\}$. In this section we consider simulation given that we know its SDF $S(\cdot)$ (for convenience, we take the sampling interval Δ_t for $\{X_t\}$ to be unity, so its Nyquist frequency f_N is $1/2$, and $S(\cdot)$ is a periodic function with a period of unity; C&E [1] for this section discusses adjusting for $\Delta_t \neq 1$). Under mild conditions (e.g., $S(\cdot)$ is square integrable), we have $\{s_{\tau}\} \longleftrightarrow S(\cdot)$, and hence knowledge of the SDF in theory gives us knowledge of the ACVS

in its entirety. This fact would seem to say that we could just use the techniques discussed in the previous section since they only require partial knowledge of the ACVS. The motivation for this section is twofold. First, there are certain SDFs that can be described simply, but their corresponding ACVSs are hard to get at. Consider $S^{(\text{PL})}(f) = |f|^{1/3}$, $|f| \leq 1/2$ (an example of a power-law SDF that might arise as a simplistic model for the increments of a certain type of ocean turbulence). While it is true that the ACVS can be described simply as

$$s_{\tau}^{(\text{PL})} = \int_{-1/2}^{1/2} |f|^{1/3} e^{i2\pi f\tau} df, \quad \tau \in \mathbb{Z},$$

a handy analytic expression for the integral is hard in general to come by, and use of numerical integration to evaluate the integral is tricky at best, particularly for large τ . Second, if the ACVS is readily available and if the required nonnegativity conditions hold, then circulant embedding is certainly the method of choice; however, if nonnegativity fails to hold, the ACVS-based alternatives to circulant embedding described in C&E [3] for Section 11.2 might be less appealing than the SDF-based approach described here.

In what follows we present an SDF-based method that closely mimics the approach to circulant embedding advocated by Dietrich and Newsam (1997). In particular the method generates a complex-valued time series from which we can extract two independent realizations of a real-valued stationary process whose statistical properties are potentially a close match for those of X_0, X_1, \dots, X_{N-1} . We refer to this method as the *Gaussian spectral synthesis method* (GSSM), where the qualifier “Gaussian” is included to emphasize that, similar to the circulant embedding method, GSSM presumes Gaussianity. References that discuss methods similar in spirit to GSSM – but differing in details – include Thompson (1973), Mitchell and McPherson (1981), Percival (1992), Chambers (1995), Davison and Hinkley (1997) and Sun and Chaika (1997).

As motivation for GSSM, consider the following connection between the circulant embedding method and the SDF $S(\cdot)$ for $\{X_t\}$.

- ▷ **Exercise [605]** Suppose that the ACVS for $\{X_t\}$ is identically zero at all lags $|\tau| \geq N$. Show that the circulant embedding weights S_k as given by Equation (601a) are such that $S_k = S(\tilde{f}_k)$. Show also that

$$\{s_{\tau} : \tau = -(N-1), \dots, N\} \longleftrightarrow \{S(\tilde{f}_k) : k = -(N-1), \dots, N\}. \quad (605a) \triangleleft$$

Hence, in this special case, we can pick off the required weights directly from the SDF for $\{X_t\}$. Since it is also true that $\{s_{\tau} : \tau \in \mathbb{Z}\} \longleftrightarrow S(\cdot)$, the inverse Fourier transform for a finite sequence and the one appropriate for an infinite sequence say that

$$s_{\tau} = \frac{1}{2N} \sum_{k=-(N-1)}^N S(\tilde{f}_k) e^{i2\pi \tilde{f}_k \tau} = \int_{-1/2}^{1/2} S(f) e^{i2\pi f\tau} df. \quad (605b)$$

In the above we can regard the middle term as a Riemann sum approximation to the integral, but an unusual one in that the approximation is actually exact (cf. the discussion following Equation (171f)). In the case where $s_{\tau} \neq 0$ for at least some $|\tau| \geq N$, the middle term is only an approximation to the integral and hence also to s_{τ} . We can obtain a better approximation to s_{τ} by decreasing the spacing between frequencies. Accordingly if we let N' be an integer greater than N , then $f'_k \stackrel{\text{def}}{=} k/(2N')$ defines a finer grid of frequencies than that of $\tilde{f}_k =$

$k/(2N)$. Assume that $S(f'_k)$ is finite for all k (this assumption does *not* hold for certain SDFs of interest – see the discussion in C&E [2]). We then have

$$s_\tau \approx \frac{1}{2N'} \sum_{k=-(N'-1)}^{N'} S(f'_k) e^{i2\pi f'_k \tau} = \frac{1}{2N'} \sum_{k=0}^{2N'-1} S(f'_k) e^{i2\pi f'_k \tau} \stackrel{\text{def}}{=} s'_\tau, \quad (606a)$$

where changing the limits over which k varies is justified because both $\{S(f'_k) : k \in \mathbb{Z}\}$ and $\{\exp(i2\pi f'_k \tau) : k \in \mathbb{Z}\}$ are periodic sequences with a period of $2N'$. Exercise [11.7] is to verify that the approximation s'_τ is real-valued. In general, since s'_τ is a Riemann sum approximation to the integral displayed in Equation (605b), for any fixed τ , the approximation s'_τ should get closer to s_τ as N' increases.

To obtain the desired simulated series, let $Z_0, Z_1, \dots, Z_{4N'-1}$ be a set of $4N'$ independent standard Gaussian RVs, and define

$$\mathcal{U}_k = \left(\frac{S(f'_k)}{2N'} \right)^{1/2} (Z_{2k} + iZ_{2k+1}), \quad 0 \leq k \leq 2N' - 1, \quad (606b)$$

and let $\{U_t\}$ be the DFT of $\{\mathcal{U}_k\}$, i.e.,

$$U_t = \sum_{k=0}^{2N'-1} \mathcal{U}_k e^{-i2\pi f'_k t} \quad (606c)$$

(note that the two equations above closely mimic Equations (601c) and (602) used by the circulant embedding method). Let $U_{\Re,t} \stackrel{\text{def}}{=} \Re\{U_t\}$ and $U_{\Im,t} \stackrel{\text{def}}{=} \Im\{U_t\}$. The burden of Exercise [11.8] is to show that $\{U_{\Re,t} : t \in \mathbb{Z}\}$ and $\{U_{\Im,t} : t \in \mathbb{Z}\}$ are Gaussian stationary processes with zero means and ACVSs given by $\{s'_\tau\}$ of Equation (606a) and that the RVs in $\{U_{\Re,t}\}$ are independent of the RVs in $\{U_{\Im,t}\}$. To the extent that s'_τ is a good approximation to s_τ for $\tau = 0, 1, \dots, N-1$, we can thus regard realizations of $U_{\Re,0}, U_{\Re,1}, \dots, U_{\Re,N-1}$ and $U_{\Im,0}, U_{\Im,1}, \dots, U_{\Im,N-1}$ as two independent realizations of processes whose ACVSs are approximately equal to that of $\{X_t\}$ at lags $|\tau| \leq N-1$ (the approximation should improve if we increase N' while holding N fixed).

What remains is to set N' . To demonstrate the interplay between N' and N , consider the AR(4) process of Equation (35a). We can compute its ACVS $\{s_\tau\}$ using the procedure described surrounding Equation (508b). The dots in all three panels of Figure 607 show s_τ versus $\tau = 0, 1, \dots, 256$. Suppose we are interested in GSSM-based simulations for sample size $N = 64$ (this is *not* the best way to generate simulations for this AR(4) process – Exercise [11.1] states the recommended procedure). If we were to set N' to $N/2 = 32$, the approximating ACVS $\{s'_\tau\}$ would be a periodic sequence with a period of 64. The solid curve in Figure 607(a) depicts this sequence. For a simulated series of length $N = 64$, we want s'_τ to be a good approximation to s_τ at lags $\tau = 0, 1, \dots, 63$ (the vertical dashed line marks lag 63). The plot indicates that s'_τ is in general a poor approximation at the desired lags. If we increase N' by setting it to $N = 64$, the approximating ACVS – the solid curve in Figure 607(b) – is now periodic with a period of 128. At the important lags, i.e., $\tau = 0, 1, \dots, 63$, the approximating s'_τ is now a better match to s_τ , but distortions are still visible. Finally let us consider setting N' to $2N = 128$. The approximating ACVS shown in Figure 607(c) is now periodic with a period of 256. The agreement between s'_τ and s_τ is visibly quite good at the lags that matter (the visible distortions at larger values of τ are of no concern for generating a series of length $N = 64$).

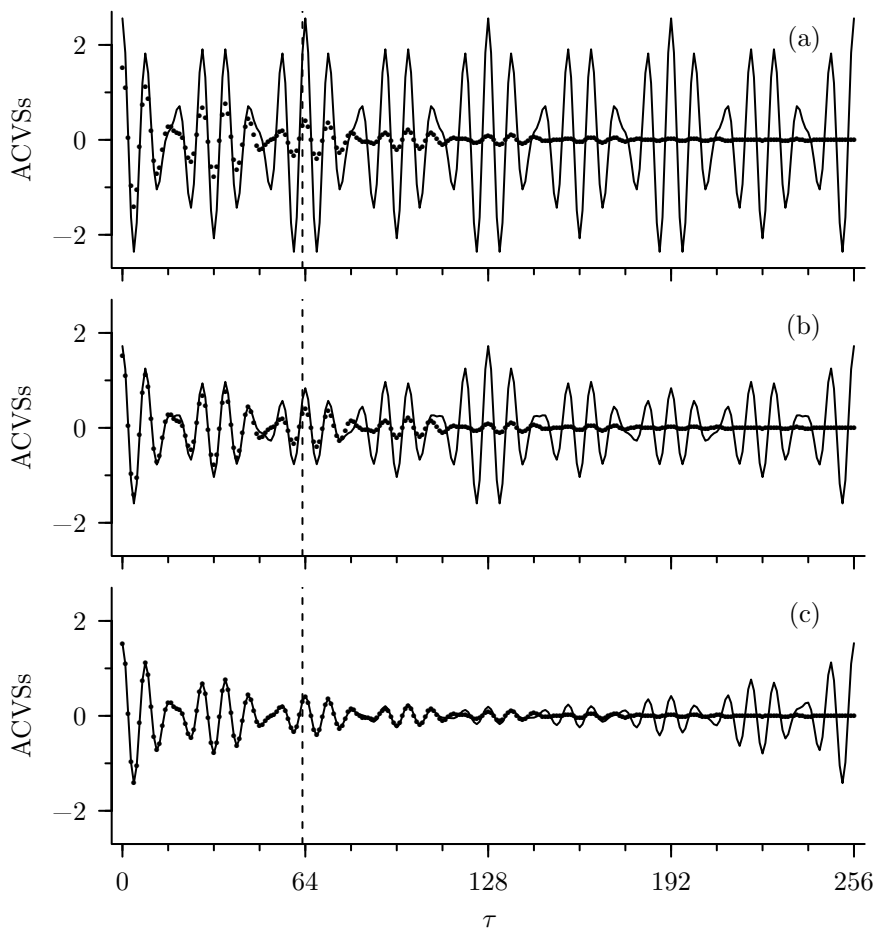


Figure 607 Actual ACVS $\{s_\tau\}$ for AR(4) process of Equation (35a) for $\tau = 0, 1, \dots, 256$ (dots in all three plots), along with three GSSM-based approximations $\{s'_\tau\}$ (solid curves). In plot (a), the approximation s'_τ is obtained via Equation (606a) with N' set to 32 and with $S(\cdot)$ set to the AR(4) SDF; plots (b) and (c) are similar, but now with N' set to 64 in the former and to 128 in the latter. The vertical dashed line marks $\tau = 63$, and the visual good agreement between s'_τ and s_τ for $|\tau| \leq 63$ in plot (c) suggests that GSSM might provide reasonably accurate simulated AR(4) series of length $N = 64$ if we set $N' = 128$.

N'	32	64	128	256	512	1024
$\sqrt{\text{NMSE}(N')}$	0.500	0.106	0.007	3.55×10^{-5}	1.81×10^{-9}	9.21×10^{-14}
$\sqrt{\text{NMSE}(N', 2N')}$	0.349	0.099	0.007	3.55×10^{-5}	1.81×10^{-9}	3.15×10^{-14}

Table 607 Square roots of $\text{NMSE}(N')$ and $\text{NMSE}(N', N'')$ criteria of Equations (608a) and (608b) as applied to AR(4) process of Equation (35a) (here $N'' = 2N'$). The values in the columns marked by $N' = 32, 64$ and 128 make use of the GSSM-based approximations s'_τ to the AR(4) ACVS shown, respectively, in Figures 607(a) to (c).

Rather than relying on a visual comparison to pick N' , we can assess various choices with the help of normalized mean square error (NMSE) criteria (the normalization essentially means that the assessment is actually done in terms of an autocorrelation sequence (ACS),

i.e., a standardized version of an ACVS). If the ACVS is known (as in our AR(4) example), the criterion for various choices of N' is

$$\text{NMSE}(N') \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\tau=0}^{N-1} \left(\frac{s'_\tau - s_\tau}{s_0} \right)^2. \quad (608a)$$

If the ACVS is not easy to get at (a motivation for using GSSM), we can consider

$$\text{NMSE}(N', N'') \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\tau=0}^{N-1} \left(\frac{s'_\tau - s''_\tau}{s''_0} \right)^2, \quad (608b)$$

where s''_τ is defined in a manner similar to s'_τ , but with $N'' > N'$ and $f''_k \stackrel{\text{def}}{=} k/(2N'')$ replacing N' and f'_k . The idea is to look at various choices of N' and N'' , e.g., $(N', N'') = (N/2, N)$, $(N, 2N)$ and so forth. Study of these NMSEs can help ascertain if increasing N' leads to enough of an improvement in the approximations to justify the resulting increase in the time that it takes a computer to generate a pair of realizations.

Table 607 looks at the square root of the two NMSE criteria for the AR(4) example (taking square roots recasts the criteria roughly as deviations between autocorrelations, which connects the criteria to quantities bounded in magnitude by unity). As expected, $\text{NMSE}(N')$ decreases as N' increases, which is consistent with what Figure 607 indicates. For $N' = 128$, an NMSE of 0.007 might suggest reasonably accurate GSSM simulations since this value is small relative to unity; on the other hand, $\text{NMSE}(256)$ is two orders of magnitude smaller, but using $N' = 256$ rather than $N' = 128$ would approximately double the amount of computer time needed to generate realizations. In the case where the true ACVS is unknown so that we must rely on $\text{NMSE}(N', N'')$ rather than $\text{NMSE}(N')$, a comparison of the two rows of Table 607 suggests that the two criteria are similar enough so that any rule of thumb for selecting N' that might work with $\text{NMSE}(N')$ would also apply to $\text{NMSE}(N', 2N')$. (When numerical precision is taken into account, the two very small entries in the column with the header $N' = 1024$ are indistinguishable from each other and from zero.)

In summary, given knowledge of the SDF $S(\cdot)$ for a Gaussian stationary process $\{X_t\}$, GSSM consists of

- [1] setting N' to an appropriate value relative to the desired N (setting N' to be less than N is not recommended – a reasonable default is often $N' = 2N$, but, in general, larger settings for N' imply more accurate GSSM-based simulations at the expense of longer computer times to generate the simulations);
- [2] computing $S(f'_k)$, $k = 0, 1, \dots, 2N' - 1$, where $f'_k = k/(2N')$;
- [3] using the $S(f'_k)$ sequence along with $4N'$ deviates from a Gaussian white noise process $\{Z_t\}$ with zero mean and unit variance to form \mathcal{U}_k as per Equation (606b); and, finally,
- [4] taking the DFT of $\{\mathcal{U}_k\}$ to obtain $U_t = U_{\Re,t} + iU_{\Im,t}$ – we can consider realizations of the real-valued sequences $U_{\Re,t}$ and $U_{\Im,t}$, $t = 0, 1, \dots, N - 1$, to be independent and – at least to some degree – to have the same statistical properties as X_t , $t = 0, 1, \dots, N - 1$.

We can compute the DFT in step [4] efficiently if $2N'$ is amenable to an FFT algorithm. Repetition of steps [3] and [4] yields as many additional pairs of realizations as desired, but a different random sample of Gaussian white noise is needed at each iteration of step [3].

As an example, let us proceed as we did in illustrating the circulant embedding method by generating realizations of length $N = 1024$ from the AR(2) process of Equation (34) (see Figure 603 and the discussion surrounding it; again we emphasize that the best way to simulate this AR(2) process is *neither* GSSM *nor* the circulant embedding method, but rather the

method described in Exercise [597]). Setting $N' = N = 1024$ allows us to directly compare the performance of GSSM and the circulant embedding method because both methods then make use of the same number of independent standard Gaussian RVs, namely, $4N = 4096$. To implement GSSM, we need to compute the SDF for the AR(2) process, namely,

$$S(f) = \frac{\sigma_2^2}{\left|1 - \sum_{j=1}^2 \phi_{2,j} e^{-i2\pi f j}\right|^2} = \frac{1}{|1 - 0.75e^{-i2\pi f} + 0.5e^{-i4\pi f}|^2}, \quad (609)$$

over the grid of frequencies $f'_k = k/2048$, $k = 0, 1, \dots, 2047$ (once we recall the assumption $\Delta_t = 1$ and once we set p to 2, the above follows from Equation (446b), which states the SDF $S(\cdot)$ for AR processes, while Equation (34) tells us that $\sigma_2^2 = 1$, $\phi_{2,1} = 0.75$ and $\phi_{2,2} = -0.5$). As noted in C&E [5] for Section 9.2, we can obtain the required sequence $\{S(f'_k)\}$ of 2048 values by taking the DFT – call it $\{G_k : k = 0, 1, \dots, 2047\}$ – of the sequence consisting of -1 , 0.75 and -0.5 followed by 2045 zeros. The desired $S(f'_k)$ is given by $1/|G_k|^2$ (see Equations (449a) and (449b), noting carefully that we need to take N' in the former to be 2048, and we need f'_k in the latter to be $k/2048$). When numerical precision is taken into account, the weight $S(f'_k)$ and the weight S_k of Equation (601a) used by the circulant embedding method are indistinguishable. As a result, a plot of $S(f'_k)$ versus k looks no different from the plot of S_k versus k shown in Figure 603(a). Additionally, if we reuse the same 4096 realizations of Gaussian white noise to form $\{\mathcal{U}_k\}$ and then $\{U_t\}$ as we did in forming $\{\mathcal{Y}_k\}$ and then $\{Y_t\}$, plots of the real and imaginary components of $\{U_t\}$ look identical to those of $\{Y_t\}$, i.e., the bottom two plots of Figure 603. For this example, GSSM generates realizations that, while in theory not exact, are virtually identical to the exact realizations obtained via circulant embedding. The degree to which we get similar results for other stationary processes is explored in Exercises [11.10] and [11.11].

Comments and Extensions to Section 11.3

[1] We have assumed for convenience that the known SDF for the stationary process $\{X_t\}$ has an associated sampling interval Δ_t of unity. If we are supplied with an SDF that is associated with a process whose sampling interval differs from unity, we cannot use this SDF with the GSSM formulation presented here (in particular the SDF for $\{X_t\}$ is periodic with a period $1/\Delta_t$, whereas our formulation presumes an SDF with unit periodicity). We need to adjust the supplied SDF to make it work with GSSM. Fortunately the adjustment is simple (the discussion here follows that of C&E [4] for Section 4.3, but the notation differs somewhat). First, take $S(\cdot)$ to be what is required by GSSM in Equation (606b) and elsewhere. Second, take $S_{\Delta_t}(\cdot)$ to be the true SDF for $\{X_t\}$ (this SDF is thus periodic with a period of $1/\Delta_t$). Given $S_{\Delta_t}(\cdot)$, we can get what our GSSM formulation needs by setting

$$S(f) = \frac{1}{\Delta_t} S_{\Delta_t} \left(\frac{f}{\Delta_t} \right).$$

Note that

$$S(f+1) = \frac{1}{\Delta_t} S_{\Delta_t} \left(\frac{f+1}{\Delta_t} \right) = \frac{1}{\Delta_t} S_{\Delta_t} \left(\frac{f}{\Delta_t} \right) = S(f),$$

and hence a periodicity of $1/\Delta_t$ for $S_{\Delta_t}(\cdot)$ implies the required unity periodicity for $S(\cdot)$. The factor of $1/\Delta_t$ just before $S_{\Delta_t}(\cdot)$ in the above is needed so that the integral of $S(\cdot)$ over an interval of unit length is equal to $s_0 = \text{var}\{X_t\}$. Thus

$$\int_{-1/2}^{1/2} S(f') df' = \frac{1}{\Delta_t} \int_{-1/2}^{1/2} S_{\Delta_t} \left(\frac{f'}{\Delta_t} \right) df' = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} S_{\Delta_t}(f) df = s_0$$

upon making the change of variable $f = f'/\Delta_t$, where, as usual, $f_{\mathcal{N}} = 1/(2\Delta_t)$ is the Nyquist frequency. Note that we have both $\{s_{\tau}\} \longleftrightarrow S_{\Delta_t}(\cdot)$ and $\{s_{\tau}\} \longleftrightarrow S(\cdot)$, which is a consequence of the

way in which Δ_t is incorporated into the Fourier transform of Equation (74a) and the inverse Fourier transform of Equation (75a) (this is consistent with the fact that the definition of the ACVS, namely, $s_\tau = \text{cov}\{X_{t+\tau}, X_t\}$, does not involve Δ_t).

After taking the GSSM realizations of $U_{\mathcal{R},0}, U_{\mathcal{R},1}, \dots, U_{\mathcal{R},N-1}$ and $U_{\mathcal{I},0}, U_{\mathcal{I},1}, \dots, U_{\mathcal{I},N-1}$ to be surrogates for realizations of X_0, X_1, \dots, X_{N-1} , we can reincorporate Δ_t by using it to associate the time index t with the actual time $t_0 + t \Delta_t$, where t_0 is the time associated with index $t = 0$.

[2] Our formulation of GSSM requires that $S(f'_k)$ be finite for $k = 0, 1, \dots, 2N' - 1$, where, as before, $f'_k = k/(2N')$. The requirement is not satisfied for all k for some stationary processes of interest. In particular, consider a power-law SDF defined by $S^{(\text{PL})}(f) = C|f|^\alpha$ for $|f| \leq 1/2$, where $-1 < \alpha < 0$. While $S^{(\text{PL})}(f'_k)$ is finite for $1 \leq k \leq 2N' - 1$, it is infinite for $k = 0$. Fortunately we can adjust GSSM to handle this problem. First, note that

$$s_0^{(\text{PL})} = \int_{-1/2}^{1/2} S^{(\text{PL})}(f) df = 2C \int_0^{1/2} f^\alpha df = \frac{2C}{\alpha+1} \left(\frac{1}{2}\right)^{\alpha+1}. \quad (610)$$

Second, as given by Equation (606a), the approximation s'_0 for $s_0^{(\text{PL})}$ is

$$s'_0 = \frac{1}{2N'} \sum_{k=0}^{2N'-1} S^{(\text{PL})}(f'_k),$$

which is infinite because $S^{(\text{PL})}(f'_0) = S^{(\text{PL})}(0) = \infty$. The adjustment consists of replacing $S^{(\text{PL})}(f'_0)$ in the above with

$$\tilde{S}_0 \stackrel{\text{def}}{=} 2N' s_0^{(\text{PL})} - \sum_{k=1}^{2N'-1} S^{(\text{PL})}(f'_k),$$

which forces the resulting expression – call it \tilde{s}'_0 – to be *exactly* equal to $s_0^{(\text{PL})}$. The adjustment that is needed to generate the simulated series is simple: just redefine \mathcal{U}_0 in Equation (606b) to be

$$\mathcal{U}_0 = \left(\frac{\tilde{S}_0}{2N'} \right)^{1/2} (Z_0 + iZ_1),$$

but keep \mathcal{U}_k for $1 \leq k \leq 2N' - 1$ as is. Note, however, that the ACVS for the simulated series is given by

$$\tilde{s}'_\tau = \frac{1}{2N'} \left(\tilde{S}_0 + \sum_{k=1}^{2N'-1} S^{(\text{PL})}(f'_k) e^{i2\pi f'_k \tau} \right)$$

rather than by s'_τ of Equation (606a) (proof of this closely follows the solution to Exercise [11.8]).

An objection here is that easy access to the process variance (as Equation (610) gives for the power-law example) violates the premise motivating this section, namely, that we have easy access to the SDF, but not the ACVS. The violation is mild in that the adjustment to GSSM assumes we have access to the zero lag component of the ACVS, but nothing more; however, if we were to multiply the power-law SDF by a continuous and even function that is bounded above and away from zero and that has unit periodicity, this newly defined SDF would still diverge to infinity as $f \rightarrow 0$. We would then have no guarantee of easy access to the process variance and might have to entertain a potentially problematic numerical integration, which would distract from using the adjusted GSSM. (There are also some stationary processes for which $S(f'_k)$ is potentially infinite at nonzero indices k , a prime example being Gegenbauer processes – see section 11.3 of Woodward et al., 2017, and references therein. Adjusting GSSM to handle such processes is tricky.)

11.4 Simulating Time Series from Nonparametric Spectral Estimates

Suppose we have a time series considered to be a realization of a portion X_0, X_1, \dots, X_{N-1} of a Gaussian stationary process with sampling time $\Delta_t = 1$ and with an SDF $S(\cdot)$ (we assume a sampling time of unity merely for convenience – C&E [1] for this section indicates what adjustments need to be made to the upcoming exposition when $\Delta_t \neq 1$). Suppose that, based upon this realization, we estimate the SDF using, say $\hat{S}(\cdot)$, and that this estimator satisfies two conditions: first, $\hat{S}(f) \geq 0$ for all f ; and second, the ACVS estimator $\{\hat{s}_\tau\}$ corresponding to $\hat{S}(\cdot)$, i.e., $\{\hat{s}_\tau\} \longleftrightarrow \hat{S}(\cdot)$, is such that $\hat{s}_\tau = 0$ for all $|\tau| \geq N$. An argument that closely parallels the proof of Exercise [605] says that

- [1] we can use the circulant embedding method to simulate a time series of length N whose *theoretical* SDF is *identical* to the estimator $\hat{S}(\cdot)$; and
- [2] the weights S_k required by that method for use in Equation (601c) satisfy $S_k = \hat{S}(\tilde{f}_k)$, where, as usual, \tilde{f}_k is defined to be $k/(2N)$.

With bootstrapping in mind, Percival and Constantine (2006) note that many nonparametric estimators satisfy the two conditions we imposed on $\hat{S}(\cdot)$. These include estimates resulting from direct SDF estimators $\hat{S}^{(D)}(\cdot)$ (see Equation (186b)); lag window estimators $\hat{S}_m^{(LW)}(\cdot)$ (Equation (248a)), but restricted to those guaranteed to be nonnegative at all frequencies; basic and weighted multitaper estimators $\hat{S}^{(MT)}(\cdot)$ and $\hat{S}^{(WMT)}(\cdot)$ (Equations (352a) and (352b)); and WOSA estimators $\hat{S}^{(WOSA)}(\cdot)$ (Equation (414b)). Hence we can use the circulant embedding method to simulate time series of length N based on estimates from almost all the nonparametric spectral estimators considered in Chapters 6, 7 and 8 (C&E [2] for this section discusses simulating series with a length different from that of the time series used to create $\hat{S}(\cdot)$, while C&E [3] looks into some nonparametric estimators for which circulant embedding is problematic).

In what follows, we match specific nonparametric SDF estimators up with the circulant embedding method. We assume $\mu = E\{X_t\}$ to be known, which means that we can take μ to be zero without loss of generality. Define

$$\tilde{X}_t = \begin{cases} X_t, & 0 \leq t \leq N-1; \\ 0, & N \leq t \leq 2N-1 \end{cases} \quad (611a)$$

(if μ is unknown, we estimate it by the sample mean $\bar{X} \stackrel{\text{def}}{=} \sum_{t=0}^{N-1} X_t/N$ and replace X_t with $X_t - \bar{X}$ in the above). Examples of simulating time series based on nonparametric SDF estimates of actual time series are given in Section 11.6.

Direct Spectral Estimators

Given a data taper $\{h_t : t = 0, 1, \dots, N-1\}$ satisfying the normalization $\sum_t h_t^2 = 1$, a direct spectral estimator is defined by

$$\hat{S}^{(D)}(f) = \left| \sum_{t=0}^{N-1} h_t X_t e^{-i2\pi f t} \right|^2$$

(this is Equation (186b) with Δ_t set to unity). Clearly we must have $\hat{S}^{(D)}(f) \geq 0$; in addition, the corresponding ACVS estimator $\{\hat{s}_\tau^{(D)}\}$ is such that $\hat{s}_\tau^{(D)} = 0$ for $|\tau| \geq N$ (see Equation (188b)), so the two conditions hold that we need for circulant embedding to work.

Define

$$\tilde{h}_t = \begin{cases} h_t, & 0 \leq t \leq N-1; \\ 0, & N \leq t \leq 2N-1. \end{cases} \quad (611b)$$

The weights S_k needed in Equation (601c) are

$$S_k = \hat{S}^{(D)}(\tilde{f}_k) = \left| \sum_{t=0}^{N-1} h_t X_t e^{-i2\pi \tilde{f}_k t} \right|^2 = \left| \sum_{t=0}^{2N-1} \tilde{h}_t \tilde{X}_t e^{-i\pi k t/N} \right|^2, \quad 0 \leq k \leq 2N-1. \quad (612a)$$

The right-hand summation is the DFT of $\{\tilde{h}_t \tilde{X}_t : t = 0, 1, \dots, 2N-1\}$ and hence can be computed efficiently if $2N$ is a length accepted by an FFT algorithm.

Lag Window Spectral Estimators

Given an ACVS estimator $\{\hat{s}_\tau^{(D)}\}$ corresponding to a direct spectral estimator and given a lag window $\{w_{m,\tau}\}$, a lag window spectral estimator is defined by

$$\hat{S}_m^{(LW)}(f) = \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau} \hat{s}_\tau^{(D)} e^{-i2\pi f \tau},$$

(this is Equation (248a) with Δ_t set to unity). As noted in Section 7.5, certain lag windows ensure that $\hat{S}_m^{(LW)}(f) \geq 0$ for all f , including the Bartlett, Daniell, Bartlett–Priestley, Parzen, Gaussian and Papoulis lag windows. For all lag window spectral estimators, the corresponding ACVS estimator $\{\hat{s}_\tau^{(LW)}\}$ is such that $\hat{s}_\tau^{(LW)} = 0$ for $|\tau| \geq N$ (see Equation (248b)). Hence the two conditions needed for circulant embedding to work hold if $\{w_{m,\tau}\}$ is selected with care.

Consider the inverse DFT of the finite sequence $\{\hat{S}^{(D)}(\tilde{f}_k)\}$ of Equation (612a):

$$\frac{1}{2N} \sum_{k=0}^{2N-1} \hat{S}^{(D)}(\tilde{f}_k) e^{i\pi k \tau/N} \stackrel{\text{def}}{=} \tilde{s}_\tau^{(D)}, \quad \tau = 0, \dots, 2N-1.$$

This inverse DFT has the following relationship to $\{\hat{s}_\tau^{(D)}\}$:

$$\tilde{s}_\tau^{(D)} = \begin{cases} \hat{s}_\tau^{(D)}, & 0 \leq \tau \leq N; \\ \hat{s}_{2N-\tau}^{(D)}, & N+1 \leq \tau \leq 2N-1 \end{cases} \quad (612b)$$

(deducible from computational details in Sections 6.7 and 7.11). Define

$$\tilde{w}_{m,\tau} = \begin{cases} w_{m,\tau}, & 0 \leq \tau \leq N; \\ w_{m,2N-\tau}, & N+1 \leq \tau \leq 2N-1 \end{cases} \quad (612c)$$

(an adaptation of Equation (314) once we recall that $w_{m,N}$ is always zero). The weights needed to use circulant embedding with lag window spectral estimates are

$$S_k = \hat{S}_m^{(LW)}(\tilde{f}_k) = \sum_{\tau=0}^{2N-1} \tilde{w}_\tau \tilde{s}_\tau^{(D)} e^{-i\pi k \tau/N}, \quad (612d)$$

i.e., the DFT of $\{\tilde{w}_\tau \tilde{s}_\tau^{(D)}\}$, which can be computed efficiently if $2N$ is compatible with an FFT algorithm.

Multitaper Spectral Estimators

Given a set of K data tapers $\{h_{k,t}\}$, $k = 0, 1, \dots, K-1$, and given a set of nonnegative weights d_k that sum to unity, a weighted multitaper estimator is defined by

$$\hat{S}^{(WMT)}(f) = \sum_{k=0}^{K-1} d_k \left| \sum_{t=0}^{N-1} h_{k,t} X_t e^{-i2\pi f t} \right|^2$$

(the above follows from Equations (352b) and (352a) with Δ_t set to unity). The condition $\hat{S}^{(\text{WMT})}(f) \geq 0$ clearly holds; in addition, the corresponding ACVS estimator is a weighted sum of K individual ACVS estimators, each corresponding to the ACVS for a direct spectral estimator and hence being zero at lags $|\tau| \geq N$, from which we can conclude that the two conditions hold that we need for circulant embedding. With $\{\tilde{h}_{k,t}\}$ defined in a manner similar to $\{\tilde{h}_t\}$ in Equation (611b), the weights needed to use circulant embedding with weighted multitaper spectral estimates follow directly from Equation (612a):

$$S_k = \hat{S}^{(\text{WMT})}(\tilde{f}_k) = \sum_{k'=0}^{K-1} d_{k'} \left| \sum_{t=0}^{2N-1} \tilde{h}_{k',t} \tilde{X}_t e^{-i\pi k t/N} \right|^2, \quad 0 \leq k \leq 2N-1. \quad (613a)$$

Weights for simulating time series whose SDFs are in agreement with a basic multitaper estimate follow from the above upon setting $d_{k'} = 1/K$.

WOSA Spectral Estimators

Given a data taper $\{h_t : t = 0, 1, \dots, N_S - 1\}$ appropriate for block size $N_S < N$, a WOSA spectral estimator is defined by

$$\hat{S}^{(\text{WOSA})}(f) = \frac{1}{N_B} \sum_{j=0}^{N_B-1} \left| \sum_{t=0}^{N_S-1} h_t X_{t+jn} e^{-i2\pi f t} \right|^2,$$

where N_B is the number of blocks to be averaged together, and n is an integer-valued shift factor satisfying the constraints of Equation (414c) (the above follows from Equations (414b) and (414a)). A minor variation on the argument used for multitaper spectral estimators says that WOSA estimators also satisfy the two conditions we need for circulant embedding to work.

Define

$$\tilde{h}'_t = \begin{cases} h_t, & 0 \leq t \leq N_S - 1; \\ 0, & N_S \leq t \leq 2N - 1. \end{cases} \quad (613b)$$

The circulant embedding weights for WOSA are

$$S_k = \hat{S}^{(\text{WOSA})}(\tilde{f}_k) = \frac{1}{N_B} \sum_{j=0}^{N_B-1} \left| \sum_{t=0}^{2N-1} \tilde{h}'_t \tilde{X}_{t+jn} e^{-i\pi k t/N} \right|^2, \quad 0 \leq k \leq 2N-1. \quad (613c)$$

Comments and Extensions to Section 11.4

[1] In generating the weights S_k needed for circulant embedding, we assumed for convenience that $\Delta_t = 1$ for the time series $\{X_t\}$ to be simulated. If not, let $\hat{S}(\cdot)$ stand for either the direct spectral estimator of Equation (186b), the lag window estimator of Equation (248a), the weighted multitaper estimator of Equation (352b) or the WOSA estimator of Equation (414b). Suppose we have computed $\hat{S}(\cdot)$ over the grid of frequencies $k/(2N\Delta_t)$, $k = 0, 1, \dots, 2N-1$. We can use these to form the required weights by setting

$$S_k = \frac{1}{\Delta_t} \hat{S}\left(\frac{k}{2N\Delta_t}\right)$$

(the right-hand side is in keeping with remarks made in C&E [1] for Section 11.3). The above weights will be *identical* to ones obtained using the right-hand sides of Equations (612a), (612d), (613a) or (613c), none of which have any dependence on Δ_t . Once a simulated series of length N has been generated using circulant embedding, we can take Δ_t into account by associating the time indices t for

the series with actual times $t_0 + t \Delta_t$, $t = 0, 1, \dots, N - 1$, where t_0 is the time associated with index $t = 0$.

[2] The procedure we have described for simulating a time series yields a simulated series of the same length N as the original time series. If we desire a simulated series of shorter length, say $N' < N$, a simple solution is to use the procedure as is and just discard the last $N - N'$ values of the original simulated series. If we are basing the simulations on a lag window estimate that uses a lag window with a truncation point m (e.g., the Bartlett lag window of Equation (269a), the Parzen window of Equation (275a) or the Papoulis window of Equation (278)), and if $m < N$, we can generate series with the shorter length $N' = m$ using weights

$$S_k = \sum_{\tau=0}^{2N'-1} \tilde{w}_\tau \tilde{s}_\tau^{(D)} e^{-i\pi k\tau/N'}, \quad 0 \leq k \leq 2N' - 1 \quad (614a)$$

(cf. Equation (612d)), for which we need to take the definitions of \tilde{w}_τ and $\tilde{s}_\tau^{(D)}$ to be

$$\tilde{w}_{m,\tau} = \begin{cases} w_{m,\tau}, & 0 \leq \tau \leq N'; \\ w_{m,2N'-\tau}, & N' + 1 \leq \tau \leq 2N' - 1 \end{cases} \quad \& \quad \tilde{s}_\tau^{(D)} = \begin{cases} \hat{s}_\tau^{(D)}, & 0 \leq \tau \leq N'; \\ \hat{s}_{2N'-\tau}^{(D)}, & N' + 1 \leq \tau \leq 2N' - 1 \end{cases}$$

(cf. Equations (612c) and (612b)). If we base the simulations on a WOSA estimate with a block size $N_S < N$, we can generate series of length $N' = N_S$ using weights

$$S_k = \frac{1}{N_B} \sum_{j=0}^{N_B-1} \left| \sum_{t=0}^{2N'-1} \tilde{h}'_t \tilde{X}_{t+jn} e^{-i\pi kt/N'} \right|^2, \quad 0 \leq k \leq 2N' - 1$$

(cf. Equation (613c)).

We can also adjust the circulant embedding method to simulate series of length $N' > N$, but with an obvious word of warning: simulating a series of length, say, $N' = 1,000,000$ based upon an SDF estimate that uses a time series with length $N = 1000$ might fail miserably in capturing the correct low-frequency properties of the phenomenon under study, particularly if, as a preprocessing step, the series has been centered by subtracting off its sample mean. Redefine

$$\tilde{X}_t = \begin{cases} X_t, & 0 \leq t \leq N - 1; \\ 0, & N \leq t \leq 2N' - 1 \end{cases} \quad \text{and} \quad \tilde{h}_t = \begin{cases} h_t, & 0 \leq t \leq N - 1; \\ 0, & N \leq t \leq 2N' - 1 \end{cases} \quad (614b)$$

(cf. Equations (611a) and (611b)). For direct spectral estimators, the weights are now

$$S_k = \left| \sum_{t=0}^{2N'-1} \tilde{h}_t \tilde{X}_t e^{-i\pi kt/N'} \right|^2, \quad 0 \leq k \leq 2N' - 1$$

(cf. Equation (612a)). The adjustments to S_k for weighted multitaper and WOSA estimators are similar: for the former, we have

$$S_k = \sum_{k'=0}^{K'-1} d_{k'} \left| \sum_{t=0}^{2N'-1} \tilde{h}_{k',t} \tilde{X}_t e^{-i\pi kt/N'} \right|^2, \quad 0 \leq k \leq 2N' - 1$$

(cf. Equation (613a)), where $\tilde{h}_{k',t}$ and \tilde{X}_t are defined in keeping with Equation (614b); for the latter, we have

$$S_k = \frac{1}{N_B} \sum_{j=0}^{N_B-1} \left| \sum_{t=0}^{2N'-1} \tilde{h}'_t \tilde{X}_{t+jn} e^{-i\pi kt/N'} \right|^2, \quad 0 \leq k \leq 2N' - 1$$

(cf. Equation (613c)), where we need to redefine

$$\tilde{h}'_t = \begin{cases} h_t, & 0 \leq t \leq N_S - 1; \\ 0, & N_S \leq t \leq 2N' - 1 \end{cases}$$

(cf. Equation (613b)) and to again use \tilde{X}_t as per Equation (614b). For lag window estimators, the weights S_k specified in Equation (614a) for the case $N' < N$ also work when $N' > N$.

[3] While the circulant embedding method is attractive for generating time series whose SDFs are in agreement with the nonparametric estimators discussed in this section, there are other nonparametric estimators for which the method is problematic. As indicated by Equation (601a), circulant embedding requires knowledge of an ACVS out to lag N in order to generate the required weights S_k . All the nonparametric estimators discussed in this section are such that the corresponding estimated ACVS, say $\{\hat{s}_\tau\}$, is known for all $\tau \in \mathbb{Z}$ (and, in particular, $\hat{s}_\tau = 0$ for $|\tau| \geq N$, which guarantees that the weights are nonnegative, as required by the circulant embedding method). For other nonparametric estimators, the corresponding estimated ACVS is either difficult to get at or ill-defined.

As an example of a nonparametric estimator whose estimated ACVS is not easily obtainable, consider nonparametric estimators that make use of a prewhitening filter such as the estimator of Equation (491b):

$$\hat{S}_X^{(\text{PC})}(f) = \frac{\hat{S}_{e,m}^{(\text{LW})}(f)}{\left| 1 - \sum_{j=1}^p \bar{\phi}_{p,j} e^{-i2\pi f j \Delta_t} \right|^2}.$$

In this example the time series $\{X_t\}$ is prewhitened using a filter with coefficients $1, -\bar{\phi}_{p,1}, \dots, -\bar{\phi}_{p,p}$; the output from the filter forms the basis for the lag window estimate $\hat{S}_{e,m}^{(\text{LW})}(\cdot)$; and this estimate is divided by the squared gain function for the filter to form the postcolored SDF estimate $\hat{S}_X^{(\text{PC})}(\cdot)$ for $\{X_t\}$. In principle, the SDF estimate $\hat{S}_X^{(\text{PC})}(\cdot)$ corresponds to the SDF for an ARMA process. Its AR coefficients are just $\bar{\phi}_{p,1}, \dots, \bar{\phi}_{p,p}$, but its MA coefficients are not directly available – they are associated with $\hat{S}_{e,m}^{(\text{LW})}(\cdot)$, and routines to extract them are numerically unstable (if we could extract them, we would have multiple ways of computing the estimated ACVS corresponding to $\hat{S}_X^{(\text{PC})}(\cdot)$; see, e.g., section 3.2.1 of Brockwell and Davis, 2016). While the estimated ACVS corresponding to $\hat{S}_{e,m}^{(\text{LW})}(\cdot)$ is readily available, it is not easy to manipulate it to get the ACVS that goes with $\hat{S}_X^{(\text{PC})}(\cdot)$. Since circulant embedding is problematic here and since it is easy to compute $\hat{S}_X^{(\text{PC})}(\cdot)$ over any desired grid of frequencies, GSSM is an attractive option for simulating time series with statistical properties closely matching those dictated by $\hat{S}_X^{(\text{PC})}(\cdot)$.

Two examples of nonparametric estimators for which the corresponding estimated ACVS is ill-defined are the adaptive multitaper estimator $\hat{S}^{(\text{AMT})}(\cdot)$ of Equation (389b) and the discretely smoothed periodogram $\hat{S}_m^{(\text{DSP})}(\cdot)$ of Equation (307). The former is defined at any particular frequency f by an iterative scheme. The corresponding estimated ACVS requires knowing $\hat{S}^{(\text{AMT})}(f)$ for all $|f| \leq f_N$, but we can only compute $\hat{S}^{(\text{AMT})}(\cdot)$ over a finite set of frequencies, which renders the ACVS ill-defined. As is also true for nonparametric estimators that make use of a prewhitening filter, GSSM is an attractive option for use with $\hat{S}^{(\text{AMT})}(\cdot)$. We noted in C&E [2] for Section 7.10 that the ACVS corresponding to $\hat{S}_m^{(\text{DSP})}(\cdot)$ does not have a standard definition; however, pending future research, circulant embedding might prove compatible with one of the ACVS definitions proposed in that C&E.

[4] Since the periodogram is a special case of a direct spectral estimator, we can use circulant embedding to simulate a Gaussian time series whose theoretical SDF is equal to the periodogram $\hat{S}^{(\text{P})}(\cdot)$ of an actual time series X_0, X_1, \dots, X_{N-1} (Equation (612a) provides the necessary weights S_k once we have set h_t in Equation (611b) to $1/\sqrt{N}$). The periodogram for the simulated series will only resemble the periodogram for the actual series to a certain extent – they will *not* be exactly equal (Exercise [11.12] demonstrates this fact). Additionally the sample mean of the simulated series in practice will differ from the sample mean of the actual time series. Keeping these two facts in mind, there is an interesting contrast between simulation via periodogram-based circulant embedding and the *method of surrogate time series* (MSTS) proposed by Theiler et al. (1992) in the context of detecting nonlinear structure in a stationary process (see also Schreiber and Schmitz, 2000; Chan and Tong, 2001; Kantz and Schreiber,

2004). Given X_0, X_1, \dots, X_{N-1} , this method starts by taking its DFT:

$$\mathcal{X}_k \stackrel{\text{def}}{=} \sum_{t=0}^{N-1} X_t e^{-i2\pi kt/N} \text{ and hence } \{X_t\} \longleftrightarrow \{\mathcal{X}_k\}$$

(see Equation (171a) with Δ_t set to unity). Note that the periodogram for $\{X_t\}$ at the k th Fourier frequency k/N is $|\mathcal{X}_k|^2/N$ (see Equation (171b)). Define ϕ_0 to be zero, and let ϕ_k , $k = 1, \dots, \lfloor (N-1)/2 \rfloor$, be IID RVs uniformly distributed over $(-\pi, \pi]$, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x . For $N/2 < k \leq N-1$, define $\phi_k = -\phi_{N-k}$. Finally, if N is even, define $\phi_{N/2}$ to be an RV that assigns a probability of $1/2$ to the points 0 and π and that is independent of the other ϕ_k RVs. The MSTS simulated series is the inverse DFT of $\{\mathcal{X}_k \exp(i\phi_k)\}$:

$$X_{\phi,t} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{X}_k e^{i\phi_k} e^{i2\pi kt/N} \text{ and hence } \{X_{\phi,t}\} \longleftrightarrow \{\mathcal{X}_k e^{i\phi_k}\}. \quad (616a)$$

The sequence $\{X_{\phi,t}\}$ is periodic with a period of N . Exercise [11.13a] is to show that it is real-valued (solving this exercise reveals that, when N is even, we need the special definition for the distribution of $\phi_{N/2}$ to ensure that $\{X_{\phi,t}\}$ is real-valued in general). Note that the periodogram for $\{X_{\phi,t}\}$ is $|\mathcal{X}_k e^{i\phi_k}|^2/N = |\mathcal{X}_k|^2/N$. Thus, for any given realization of the ϕ_k RVs, the simulated series $X_{\phi,0}, X_{\phi,1}, \dots, X_{\phi,N-1}$ has *identically* the same periodogram at the Fourier frequencies as X_0, X_1, \dots, X_{N-1} ; moreover, their sample means are the same since $\sum_{t=0}^{N-1} X_t = \mathcal{X}_0 = \mathcal{X}_0 e^{i\phi_0} = \sum_{t=0}^{N-1} X_{\phi,t}$. By contrast, a series simulated via circulant embedding has a periodogram and sample mean differing in practice from those of $\{X_t\}$.

As noted in the solution to Exercise [602], the stationary process from which the simulated series from circulant embedding arise is a harmonic process formulated as per Equation (35c), which involves linear combinations of sines and cosines with random amplitudes (this is the defining equation for a harmonic process). For MSTS, with the time series $\{X_t\}$ regarded as fixed and with φ_k defined to be such that $\mathcal{X}_k = |\mathcal{X}_k| \exp(i\varphi_k)$, we have

$$X_{\phi,t} = \bar{X} + \sum_{k=1}^{\lfloor N/2 \rfloor} D_k \cos(2\pi kt/N + \phi'_k), \text{ where } D_k \stackrel{\text{def}}{=} \begin{cases} 2|\mathcal{X}_k|/N, & 1 \leq k < N/2; \\ |\mathcal{X}_{N/2}|/N, & k = N/2, \end{cases} \quad (616b)$$

and

$$\phi'_k \stackrel{\text{def}}{=} \begin{cases} \varphi_k + \phi_k + 2\pi, & \varphi_k + \phi_k \leq -\pi; \\ \varphi_k + \phi_k - 2\pi, & \varphi_k + \phi_k > \pi; \\ \varphi_k + \phi_k, & \text{otherwise,} \end{cases} \quad (616c)$$

with the RVs ϕ'_k having the same multivariate distribution as the RVs ϕ_k . Exercise [11.13b] is to verify the above, while [11.13c] is to establish that, when N is odd, the process $\{X_{\phi,t}\}$ is a harmonic process formulated as per Equation (35d), but one involving a linear combination of cosines with random phases and fixed amplitudes; on the other hand, when N is even, $\{X_{\phi,t}\}$ is in general *not* a harmonic process. Whereas the realizations from circulant embedding are from a Gaussian distribution because we specified the random amplitudes to be Gaussian RVs, MSTS realizations do not necessarily obey a Gaussian distribution exactly, but, due to a central limit effect, the sum of $N-1$ randomly phase cosines will be close to Gaussian if the D_k amplitudes are not dominated by a few large values (Walden and Prescott, 1983; Sun and Chaika, 1997). The relative merits of circulant embedding and MSTS depend on the rationale for simulating time series, but certainly circulant embedding has an advantage over MSTS in any applications for which variations in the periodogram or in the sample mean are desirable.

11.5 Simulating Time Series from Parametric Spectral Estimates

Given a time series considered to be a realization of a portion X_0, X_1, \dots, X_{N-1} of the zero mean p th order causal (and hence stationary) AR process

$$X_t = \sum_{j=1}^p \phi_{p,j} X_{t-j} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a zero mean white noise process, suppose we estimate the AR parameters $\phi_{p,1}, \dots, \phi_{p,p}$ and $\sigma_p^2 = \text{var}\{\epsilon_t\}$ using one of the methods described in Chapter 9 (Yule–Walker, Burg, least squares or maximum likelihood). Let $\hat{\phi}_{p,1}, \dots, \hat{\phi}_{p,p}$ and $\hat{\sigma}_p^2$ denote these parameter estimates. We can form a corresponding parametric SDF estimate, say $\hat{S}(\cdot)$, by plugging the parameter estimates into the functional form for the SDF for an AR(p) process:

$$\hat{S}(f) = \frac{\hat{\sigma}_p^2}{\left| 1 - \sum_{j=1}^p \hat{\phi}_{p,j} e^{-i2\pi f j} \right|^2} \quad (617)$$

(see Equation (446b) with the sampling time Δ_t taken to be unity, as assumed elsewhere in this chapter). Our task is to generate realizations from the estimated AR(p) process.

Suppose that the AR coefficient estimates $\hat{\phi}_{p,1}, \dots, \hat{\phi}_{p,p}$ correspond to those of a causal AR process (causality is guaranteed for Yule–Walker estimates and always happens in practice for Burg estimates, but not necessarily for least squares estimates). If we are willing to assume Gaussianity, we can readily adapt the simulation scheme of Equation (597a) to get the desired realizations of the estimated AR(p) process. This scheme depends just on the $p+1$ parameters $\phi_{p,1}, \dots, \phi_{p,p}$ and σ_p^2 , and the adaptation replaces these with $\hat{\phi}_{p,1}, \dots, \hat{\phi}_{p,p}$ and $\hat{\sigma}_p^2$. In particular, use of these estimates in conjunction with Equations (597b) and (596a) yields everything needed in Equation (597a) to generate a simulated AR series of length N based on a random sample $\{Z_t\}$ of size N from a standard Gaussian distribution (to simulate a series of length $N' \neq N$, take $\{Z_t\}$ to be a random sample of size N' instead).

We can also check the Gaussian assumption by forming the observed prediction errors in keeping with Equation (596b), but some adjustments are needed to put the errors on a common footing. Recall that the variance for the unobservable predictions error $\bar{\epsilon}_t(t)$, $t = 0, 1, \dots, p-1$, is σ_t^2 so that $\bar{\epsilon}_t(t)/\sigma_t$ has unit variance; additionally, $\bar{\epsilon}_t(t)/\sigma_p$ has unit variance for $t = p, \dots, N-1$. Equation (596a) allows us to express σ_t^2 for a given $t < p$ in terms of σ_p^2 and $\phi_{t+1,t+1}, \dots, \phi_{p,p}$. In particular

$$\begin{aligned} \sigma_{p-1}^2 &= \frac{\sigma_p^2}{1 - \phi_{p,p}^2} \\ \sigma_{p-2}^2 &= \frac{\sigma_{p-1}^2}{1 - \phi_{p-1,p-1}^2} = \frac{\sigma_p^2}{\prod_{j=p-1}^p (1 - \phi_{j,j}^2)} \quad (\text{not needed if } p = 1) \\ &\vdots \\ \sigma_0^2 &= \frac{\sigma_1^2}{1 - \phi_{1,1}^2} = \frac{\sigma_p^2}{\prod_{j=1}^p (1 - \phi_{j,j}^2)} \quad (\text{not needed if } p = 1 \text{ or } 2). \end{aligned}$$

For $t = 0, \dots, p-1$, we thus have

$$\sigma_t^2 = \frac{\sigma_p^2}{\prod_{j=t+1}^p (1 - \phi_{j,j}^2)}.$$

The observed prediction errors are defined in a manner similar to the unobservable prediction errors of Equation (596b):

$$\begin{aligned}\vec{e}_0(0) &= X_0 \\ \vec{e}_t(t) &= X_t - \sum_{j=1}^t \hat{\phi}_{t,j} X_{t-j}, \quad t = 1, \dots, p-1 \quad (\text{not needed if } p = 1) \\ \vec{e}_t(t) &= X_t - \sum_{j=1}^p \hat{\phi}_{p,j} X_{t-j}, \quad t = p, \dots, N-1.\end{aligned}\quad (618a)$$

The observed prediction errors normalized so that their variances are on a common footing are

$$\vec{Z}_t(t) \stackrel{\text{def}}{=} \frac{\vec{e}_t(t)}{\hat{\sigma}_t}, \quad \text{where } \hat{\sigma}_t^2 \stackrel{\text{def}}{=} \begin{cases} \hat{\sigma}_p^2 / \prod_{j=t+1}^p (1 - \hat{\phi}_{j,j}^2), & 0 \leq t \leq p-1; \\ \hat{\sigma}_p^2, & p \leq t \leq N-1. \end{cases} \quad (618b)$$

If $\{X_t\}$ is indeed a Gaussian AR(p) process, these normalized errors $\{\vec{Z}_t(t)\}$ should resemble a random sample from a standard Gaussian distribution (they will not be exactly so due to the use of estimated AR parameters). In the spirit of bootstrapping (see, e.g., Efron and Gong, 1983; Swanepoel and Van Wyk, 1986; and Davison and Hinkley, 1997), rather than creating simulated series via Equation (597a) using a random sample of size N from a standard Gaussian distribution, we can use a random sample (with replacement) from $\{\vec{Z}_t(t) : t = 0, 1, \dots, N-1\}$ (this scheme is used with the ocean wave data in Section 11.6).

If the AR coefficient estimates do *not* correspond to those of a causal AR process (as is a possibility with least squares estimators), this fact will become apparent either because $|\hat{\phi}_{p,p}| \geq 1$ or because, when attempting to use the coefficients estimates $\hat{\phi}_{p,j}$ to get corresponding coefficients for the best linear predictors of orders less than p via recursive use of Equation (597b), the estimate $\hat{\phi}_{j,j}$ for some $\phi_{j,j}$, $j < p$, is such that $|\hat{\phi}_{j,j}| \geq 1$ (see C&E [5] for Section 9.4). Since we cannot estimate all the coefficients required by the simulation scheme of Equation (597a), we cannot use that method to generate realizations exactly consistent with the statistical properties of the estimated AR(p) process. If, however, the coefficients $\hat{\phi}_{p,j}$ correspond to a stationary acausal AR(p) process, then the corresponding SDF estimate $\hat{S}(\cdot)$ of Equation (617) has all the properties of an SDF (see C&E [1] for Section 9.2; note that, if the noncausal coefficients $\hat{\phi}_{p,j}$ were to correspond to a nonstationarity process, then $\hat{S}(f) = \infty$ for some f since one of the roots of $1 - \sum_{j=1}^p \hat{\phi}_{p,j} z^{-j}$ would lie on the unit circle). In lieu of an exact simulation method and in view of the fact that we can readily compute $\hat{S}(\cdot)$ over various grids of frequencies (see C&E [5] for Section 9.2), a good candidate for an approximate simulation method is GSSM – we just need to substitute $\hat{S}(f'_k)$ for $S(f'_k)$ in Equation (606b).

Comments and Extensions to Section 11.5

[1] We can regard the observed prediction errors $\{\vec{e}_t(t)\}$ of Equation (618a) as forward errors in the same spirit as ones encountered in Burg's algorithm (see Equation (467a), which is the same as Equation (618a) when $p \leq t \leq N-1$). Burg's algorithm also makes use of observed backward prediction errors. We can also entertain these here – denote them by $\{\overleftarrow{e}_t(t) : t = 0, 1, \dots, N-1\}$. A simple way to define the backward errors is take the right-hand sides of the equalities in Equation (618a) and replace X_0, X_1, \dots, X_{N-1} with its time reversal, i.e., $X_{N-1}, X_{N-2}, \dots, X_0$. The observed forward and backward prediction errors do not constitute a set of $2N$ uncorrelated deviates, but, nonetheless, we can entertain bootstrapping schemes in which we randomly sample (with replacement) from $\vec{Z}_0(0), \vec{Z}_0(0), \vec{Z}_1(1), \vec{Z}_1(1), \dots, \vec{Z}_{N-1}(N-1), \vec{Z}_{N-1}(N-1)$, where $\vec{Z}_t(t)$ is defined by replacing $\vec{e}_t(t)$ in the definition for $\vec{Z}_t(t)$ in Equation (618b) with $\overleftarrow{e}_t(t)$.

11.6 Examples of Simulation of Time Series

Ocean Wave Data

We have previously considered the ocean wave data in Sections 6.8, 7.12, 8.9, 9.12 and 10.15. This series $\{X_t\}$ is shown in Figure 620(a), which is a copy of Figure 225. The series is of length $N = 1024$, has a sampling interval of $\Delta_t = 1/4$ sec and hence spans a total of 256 sec. Let us first use these data as an example of simulating a time series whose properties are in keeping with a lag window spectral estimate. The dark curves in Figures 318 and 319 show several lag window SDF estimates for this series, all of which are smoothed versions of a direct spectral estimate $\hat{S}^{(D)}(\cdot)$ using an $NW = 2/\Delta_t$ Slepian data taper (the light curves in the figures show $\hat{S}^{(D)}(\cdot)$). Tapering is advised because, as demonstrated in Figure 226, the periodogram shows evidence of bias due to leakage. The $m = 150$ Parzen lag window estimate in Figure 318 tracks $\hat{S}^{(D)}(\cdot)$ nicely and is considerably smoother, but arguably too bumpy over $0.2 \text{ Hz} \leq f \leq 1.0 \text{ Hz}$ in light of a hypothesized monotonic rolloff. The heavier degree of smoothing achieved by the four lag window estimates shown in Figure 319 depicts the rolloff better, but at the price of smearing out the dominant peak centered near $f = 0.16 \text{ Hz}$ (the $m = 150$ Parzen estimate does a better job of capturing this peak). These four lag window estimates have comparable smoothing window bandwidths, and, in our discussion of that figure, we settled on the $m = 23.666$ Gaussian lag window estimate (the dark curve in Figure 319(d)) as the most attractive of the four. Under the assumption that the ocean wave series is a realization of a stationary process, the $m = 150$ Parzen and Gaussian lag window SDF estimates are good candidates for simulating time series whose statistical properties are in agreement with those of the actual series.

Plots (b) to (e) in Figure 620 show four simulations of $\{X_t\}$ via the circulant embedding method described in Section 11.2 and adapted in Section 11.4 for use with lag window estimates. Plots (b) and (d) show simulations based on the $m = 150$ Parzen estimate. The weights S_k needed by circulant embedding are calculated using Equation (612d), which, in forming $\tilde{w}_{m,\tau}$ as per Equation (612c), makes use of the Parzen lag window $w_{m,\tau}$ of Equation (275a) with m set to 150. The weights also need $\hat{s}_\tau^{(D)}$, which is formed as per Equation (612b) and which in turn depends on the ACVS estimate $\hat{s}_\tau^{(D)}$ corresponding to the direct spectral estimate. Equation (188b) defines $\hat{s}_\tau^{(D)}$, but, prior to using that equation, we must center the time series by subtracting off its sample mean $\bar{X} \doteq 209.1$ because the ocean wave data are not compatible with the assumed stationary process having a mean of zero; in addition, the data taper used in Equation (188b) is the Slepian taper as approximated by Equation (196b) with W set to $2/N$. As must be true, all the weights S_k are nonnegative, with the smallest and largest being 0.4141 and 4.261×10^6 (these are associated with indices $k = 810$ and $k = 81$ and with frequencies 1.582 Hz and 0.1582 Hz). The simulations in Figures 620(b) and (d) are first $N = 1024$ parts of the real and imaginary components of the complex-valued series $\{Y_t\}$ of length $2N = 2048$ – see Equation (602), which, in view of Equation (601c), depends upon the weights S_k and a random sample of size $4N = 4096$ from a standard Gaussian distribution. The actual series has a prominent nonzero sample mean, whereas the simulated series are drawn from a zero mean stationary process (but their sample means are different from zero). To compensate for this mismatch, we have added the sample mean $\bar{X} \doteq 209.1$ to both simulated series. This addition results in the sample mean for the series in plot (b) being 230.6 , and it is 205.8 in plot (d). Visually comparing the two simulated series with the actual series in Figure 620(a) does not suggest a serious mismatch in their statistical properties.

Plots (c) and (e) in Figure 620 show simulations created in a manner similar to those in (b) and (d), but now using the Gaussian lag window estimate rather than the $m = 150$ Parzen. Some details are necessarily different. We now use the Gaussian lag window of Equation (276) with m set to 23.666 . The smallest and largest weights S_k are 0.5272 and 2.039×10^6 (these are associated with indices $k = 809$ and $k = 81$ and with frequencies

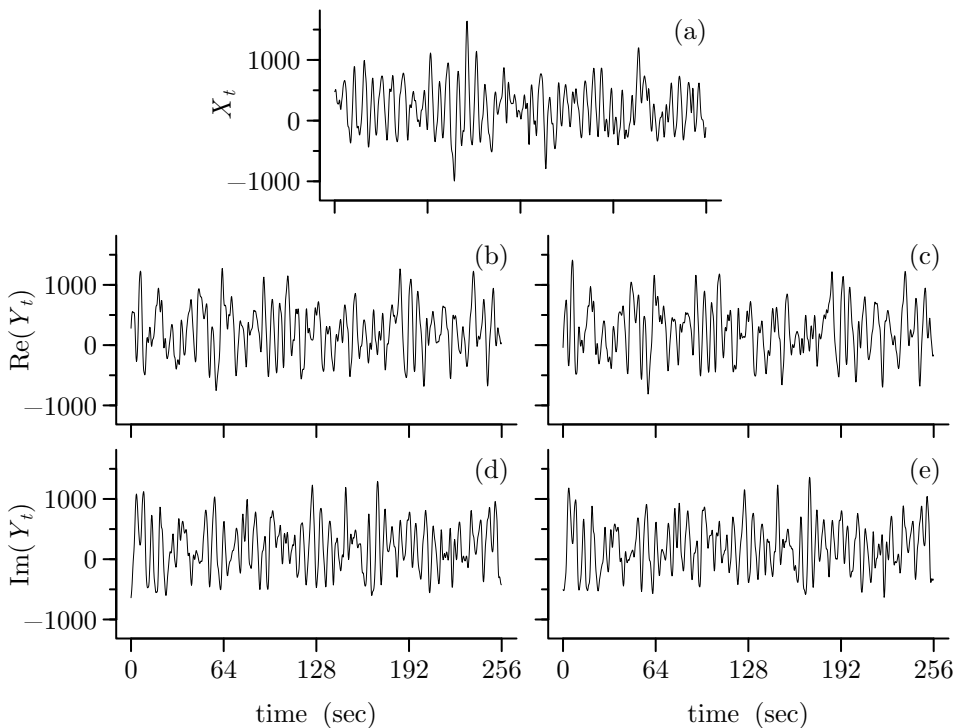


Figure 620 Ocean wave data $\{X_t : t = 0, 1, \dots, N - 1\}$ (plot (a) – also shown in Figure 225) and four simulations of this series based upon circulant embedding (plots (b) to (e)). Here $N = 1024$, and the series has a sampling interval of $\Delta_t = 1/4$ sec, resulting in a time span of 256 sec for it and its simulations. For plots (b) and (d), the weights S_k , $k = 0, 1, \dots, 2N - 1$, used to create the complex-valued series $\{Y_t : t = 0, 1, \dots, 2N - 1\}$ of Equation (602) are based on the $m = 150$ Parzen lag window estimate shown as the dark curve in Figure 318; (b) and (d) show the real and imaginary parts of Y_t , $t = 0, 1, \dots, N - 1$. Plots (c) and (e) are similar, but now use the $m = 23.666$ Gaussian lag window estimate shown as the dark curve in Figure 319(d). The same sample of $4N$ deviates from a standard Gaussian distribution is used to create the Y_t series from both lag window estimates.

1.580 Hz and 0.1582 Hz). The sample means for the series in plots (c) and (e) are 241.7 and 199.8. In addition we used *exactly* the same sample of size $4N = 4096$ from a standard Gaussian distribution as in the Parzen case. The simulations in plots (b) and (c) are thus *not* independent (in fact they are highly correlated – their sample correlation coefficient $\hat{\rho}$ is 0.96). The same is true for plots (d) and (e) (now $\hat{\rho} \doteq 0.97$). Using the same random sample allows us to visualize changes in the simulated series due entirely to the different weighting schemes (Parzen and Gaussian). Simulations (b) and (c) track each other well, as do (d) and (e). The two weighting schemes thus do not yield simulated series with much visual difference.

In view of Figure 620, should we use the $m = 150$ Parzen or the Gaussian lag window estimate as the basis for simulating the ocean wave data? If our interest is in simulating series whose spectral properties over $0.2 \text{ Hz} \leq f \leq 1.0 \text{ Hz}$ are estimated as best as possible under the working hypothesis of a monotonic rolloff, the discussion surrounding Figures 318 and 319 suggests using the Gaussian lag window estimate; on the other hand, oceanographic considerations might support the hypothesis that, due to the mechanism generating ocean waves, the high-power part of the SDF has a peak centered near 0.16 Hz and that the peak has a narrower bandwidth than what the Gaussian lag window estimate suggests. This hypothesis would favor the $m = 150$ Parzen estimate if we want the simulated series to be faithful to the high-power part of the SDF. The choice between the Parzen or the Gaussian estimate thus

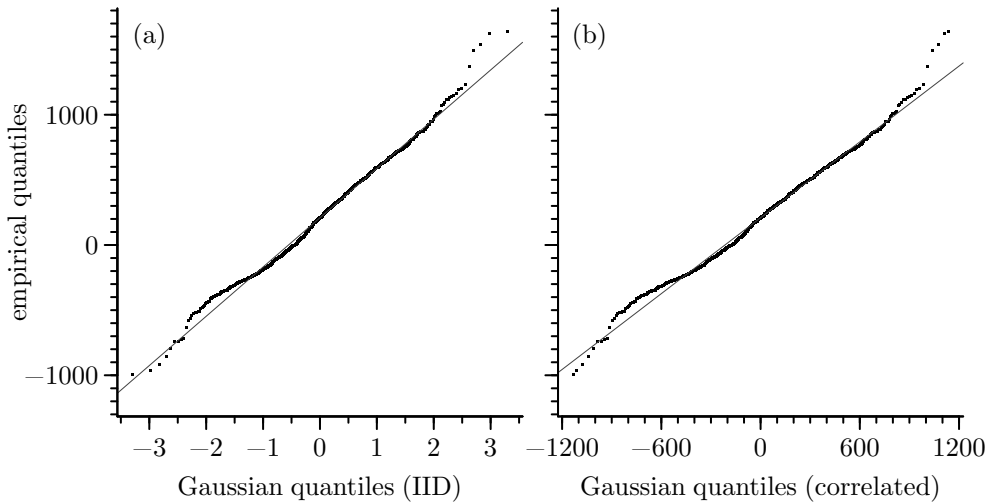


Figure 621 Gaussian QQ plots for ocean wave data (Figure 620(a) shows the series itself). The empirical quantiles (order statistics) for the time series are plotted as small dots versus the expected value of quantiles under an IID Gaussian assumption (left-hand plot) and under an assumption of a Gaussian stationary process with an SDF given by an $m = 150$ Parzen lag window estimate $\hat{S}_m^{(LW)}(\cdot)$ (right-hand plot; the dark curve in Figure 318 shows $\hat{S}_m^{(LW)}(\cdot)$). The lines are simple linear least squares fits of the empirical versus theoretical quantiles.

depends upon what we want to achieve with the simulated series in terms of delving into various hypotheses about the ocean wave data. (If we were willing to entertain the hypothesis that suggests the Gaussian estimate in addition to the hypothesis that suggests Parzen, we could entertain a scheme that would somehow merge the weights generated by the two estimates to yield simulations in support of both hypotheses).

Theoretical justification for all four simulations in Figure 620 follows if the ocean wave data obey a Gaussian distribution. To investigate the univariate distribution for $\{X_t\}$, Figure 621(a) shows a Gaussian QQ plot for this series constructed in the following manner (section 1.6, Brockwell and Davis, 2016; see also Chambers et al., 1983). Let $X_{(t)}$, $t = 0, 1, \dots, N - 1$, denote the order statistics for the series, i.e., the X_t values rearranged from smallest to largest so that $X_{(0)} \leq X_{(1)} \leq \dots \leq X_{(N-1)}$. The small dots in Figure 621(a) show $X_{(t)}$ versus q_t , where $q_t \stackrel{\text{def}}{=} \Phi^{-1}([t + 0.5]/N)$, and $\Phi^{-1}(p)$ is the $p \times 100\%$ percentage point of the standard Gaussian distribution; i.e., we are plotting empirical quantiles versus corresponding theoretical quantiles. Visual evidence against Gaussianity consists of $X_{(t)}$ not appearing to vary linearly with q_t . The line in Figure 621(a) is a least squares fit of the linear model $E\{X_{(t)}\} = \alpha + \beta q_t$, where α and β are relatable to μ and σ , the theoretical mean and standard deviation of the ocean wave series under the null hypothesis of Gaussianity. The bulk of the dots seem to have a linear alignment, but there are notable deviations in the lower and upper tails. To quantify the degree of linearity, consider the coefficient of determination

$$R^2 \stackrel{\text{def}}{=} \frac{\left(\sum_{t=0}^{N-1} (X_{(t)} - \bar{X}) q_t\right)^2}{\sum_{t=0}^{N-1} (X_{(t)} - \bar{X})^2 \sum_{t=0}^{N-1} q_t^2}, \quad (621)$$

a standard measure for assessing the efficacy of a simple linear regression (see, e.g., Weisberg, 2014). This coefficient can range from 0 to 1. If R^2 is too small, there is evidence against the null hypothesis. To determine what constitutes “too small,” we need the distribution of R^2 under the null hypothesis. This distribution is analytically intractable, but we can approximate

it by generating a large number, say $N_R = 100,000$, of simulated ocean wave series using circulant embedding based on the $m = 150$ Parzen lag window estimate (if we use the Gaussian lag window estimate instead, the results stated in what follows are virtually identical). The simulated series are each constructed in the same manner as those shown in parts (b) and (d) of Figure 620, but with each set of two series being formed using a different random sample of size $4N$ from a standard Gaussian distribution. The simulated series are necessarily Gaussian, and, for each of them, we can compute R^2 , which leads to an empirical approximation to its distribution under the null hypothesis of a zero mean stationary process whose SDF is the $m = 150$ Parzen lag window estimate. For the N_R simulated series, R^2 ranges from 0.96349 to 0.99948. For the ocean wave series, $R^2 \doteq 0.99227$, which is smaller than that of 88,406 of the 100,000 simulated series. The p -value, i.e., the observed significance level, is 0.116, which means we would fail to reject the null hypothesis at a level of significance $\alpha = 0.1$. Evidence against Gaussianity is not particularly strong.

A critique of the QQ plot shown in Figure 621(a) is that the theoretical quantiles q_t are approximations to the expected values of the quantiles for N IID RVs obeying a standard Gaussian distribution; i.e., there is an assumption of uncorrelatedness in constructing the QQ plot, whereas the time series we are dealing with is associated with RVs having a covariance structure that should be in keeping with the $m = 150$ Parzen lag window estimate $\hat{S}_m^{(LW)}(\cdot)$. This suggests plotting the empirical quantiles against the expected values of the quantiles from a time series of length N whose SDF is given by $\hat{S}_m^{(LW)}(\cdot)$. We can approximate these expected values by considering again the N_R simulated series. For each pair $\{Y_{\mathcal{R},t}\}$ and $\{Y_{\mathcal{S},t}\}$ of simulated series of length N , we form their order statistics $\{Y_{\mathcal{R},(t)}\}$ and $\{Y_{\mathcal{S},(t)}\}$. Separately for each t , we average the N_R order statistics collected from the $N_R/2$ individual realizations of $Y_{\mathcal{R},(t)}$ and $Y_{\mathcal{S},(t)}$ – let q'_t denote this average, and regard it as an approximation to a theoretical replacement for q_t that takes into account the statistical properties dictated by $\hat{S}_m^{(LW)}(\cdot)$. Figure 621(b) shows another Gaussian QQ plot for the ocean wave series, but this time consisting of the empirical quantiles $X_{(t)}$ plotted against the theoretical quantiles q'_t . The degree of linearity in the two QQ plots in Figure 621 appears remarkably similar. With q'_t replacing q_t in Equation (621), we now have $R^2 \doteq 0.99144$ in comparison to 0.99227 for the QQ plot in Figure 621(a). To assess if R^2 is too small, we use a procedure analogous to the one previously described. For the N_R simulated series, R^2 now ranges from 0.96119 to 0.99965. For the ocean wave series, this R^2 is now smaller than that for 92,288 of the N_R simulated series, yielding a p -value of 0.077, which means we would fail to reject the null hypothesis at a level of significance $\alpha = 0.05$, but would reject for $\alpha = 0.1$. There is thus some evidence against Gaussianity, but not particularly strong. Since the evidence against the ocean wave data being Gaussian is somewhat weak, there is no compelling reason to be dissatisfied with the simulations shown in Figure 620.

Let us now consider using simulated series to assess the variability in the Gaussian lag window estimate $\hat{S}_m^{(LW)}(\cdot)$ shown in Figure 319(d) and redisplayed as the solid curve in Figure 623a. The assessment involves a bootstrapping scheme advocated in Percival and Constantine (2006), which is based on the same circulant embedding method that led to the simulations of the ocean wave series shown in Figures 620(c) and (e). The scheme starts by generating a large number N_R of simulated series based upon $\hat{S}_m^{(LW)}(\cdot)$ (as before, we take “large” to be 100,000). For each such series, we compute a Gaussian lag window estimate formed in exactly the same manner as we formed $\hat{S}_m^{(LW)}(\cdot)$ (in particular, we use an $NW = 2/\Delta_t$ Slepian data taper after centering the simulated series by subtracting off its sample mean, and we set the lag window parameter m to 23.666). Let $\tilde{S}_n^{(LW)}(\tilde{f}_k)$ be the lag window estimate at frequency $k/(2N\Delta_t)$ for the n th simulated series, $n = 1, 2, \dots, N_R$. For each frequency, we sort the N_R estimates $\tilde{S}_n^{(LW)}(\tilde{f}_k)$ – call these $\tilde{S}_{(n)}^{(LW)}(\tilde{f}_k)$ so that

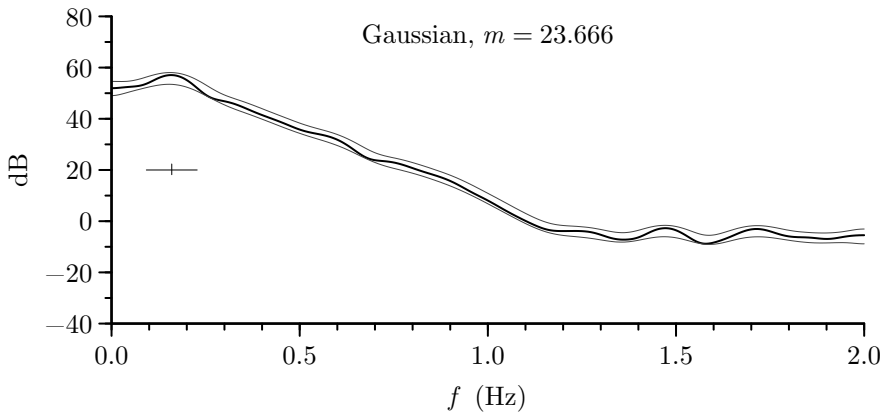


Figure 623a Gaussian lag window spectral estimate $\hat{S}_m^{(LW)}(\cdot)$ with $m = 23.666$ for ocean wave data (dark curve, reproduced from Figure 319(d)). The width and height of the crisscross gives the bandwidth measure $B_{\mathcal{U}}$ (Equation (256a)) and the length of a 95% CI for $10 \log_{10} S(f)$ (Equation (266a)). The thin curves surrounding the dark curve show bootstrapped 95% CIs based upon circulant embedding.

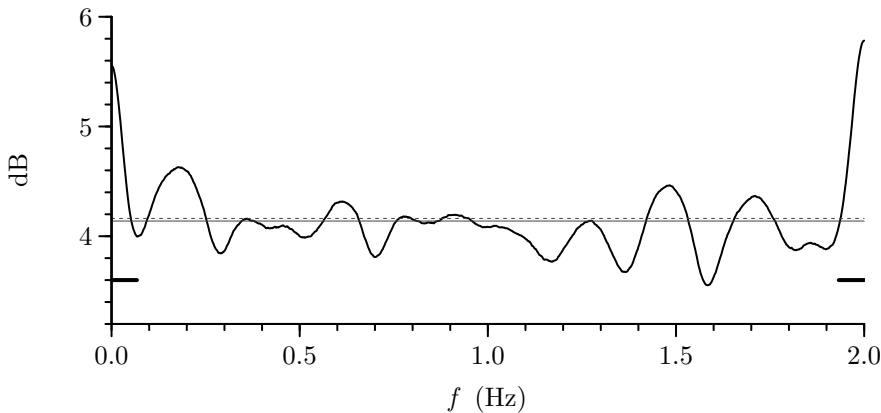


Figure 623b Widths of bootstrapped 95% CIs shown as thin curves in Figure 623a (dark curve). The solid and dashed thin horizontal lines show, respectively, the average of these dB widths (4.14 dB) and the length of a 95% CI as calculated by Equation (266a) (4.16 dB; the height of the crisscross in Figure 623a also depicts this length). The widths of the short thick lines emanating from $f = 0$ Hz and $f = 2$ Hz depict the half-bandwidth measure $B_{\mathcal{U}}/2$ as calculated by Equation (256a) (the width of the crisscross in Figure 623a depicts $B_{\mathcal{U}}$).

$\tilde{S}_{(n)}^{(LW)}(\tilde{f}_k) \leq \tilde{S}_{(n+1)}^{(LW)}(\tilde{f}_k)$ for $n = 1, 2, \dots, N_R - 1$. We then pick out $\tilde{S}_{(\lfloor 0.975 N_R \rfloor)}^{(LW)}(\tilde{f}_k)$ and $\tilde{S}_{(\lfloor 0.025 N_R \rfloor)}^{(LW)}(\tilde{f}_k)$ to serve as upper and lower bounds for a bootstrapped pointwise 95% CI for the unknown $S(\tilde{f}_k)$. These bounds are shown as the two thin curves in Figure 623a, where they depict a 95% CI for $10 \log_{10} S(\tilde{f}_k)$. The SDF estimate $\hat{S}_m^{(LW)}(\cdot)$ is always trapped between the upper and lower bounds of the bootstrapped CIs, but there are some frequencies where $\hat{S}_m^{(LW)}(\tilde{f}_k)$ is just barely below the upper bound, and others where it is barely above the lower. The crisscross in Figure 623a is redisplayed from Figure 319(d). Its height gives the length of a 95% CI based upon Equation (266a). This height appears to be comparable to the typical widths of the bootstrapped CIs. Figure 623b compares the bootstrapped CIs with the theory-based CIs in more detail. The dark curve shows the widths of the bootstrapped CIs versus frequency. The average of these dB widths is 4.14 dB, which is indicated by a solid thin line. The length of the theory-based 95% CI is 4.16 dB and is indicated by a dashed thin

line (barely distinguishable from the solid line). The two methods for generating CIs for the unknown SDF yield remarkably similar results.

Figure 623b shows the widths of the bootstrapped CIs being elevated near the zero and Nyquist frequencies. This is not particularly surprising since the variance of the direct spectral estimate $\hat{S}^{(D)}(\cdot)$ that is being smoothed to produce $\hat{S}_m^{(LW)}(\cdot)$ increases at zero and Nyquist frequencies (see Section 6.6). This fact would imply an increase in the variability of $\hat{S}_m^{(LW)}(\cdot)$ as we approach zero and Nyquist frequencies, which would translate into an increase in CIs near these frequencies. Evoking the bandwidth B_U of Equation (256a), the half-bandwidth $B_U/2$ of $\hat{S}_m^{(LW)}(\cdot)$ comes into play here in that we might expect $\hat{S}_m^{(LW)}(\cdot)$ at frequencies in the intervals $[0, B_U/2]$ and $[f_N - B_U/2, f_N]$ to be subject to increased variability. The short thick lines in Figure 623b depict these two intervals. The regions of elevated widths fall within these intervals, but the intervals fail to precisely delineate the regions of elevation.

As a final example of how simulations can help with the analysis of the ocean wave series, let us reconsider its Burg AR(25) SDF estimate $\hat{S}(\cdot)$ (shown as the thick curve in Figure 625a, which is the same as the thick curves in both panels of Figure 500). Our goal is to assess the variability in $\hat{S}(\cdot)$. To do so, we adapt the simulation scheme discussed in Section 11.5. The adaptation yields the bootstrapping scheme proposed by Swanepoel and Van Wyk (1986), but with two notable modifications. First, their scheme uses the normalized observed prediction errors $\overline{Z}_t(t)$ of Equation (618b), but only those indexed by $t = p, p + 1, \dots, N - 1$, whereas we use all N such errors (for this example, $p = 25$ and $N = 1024$). Second, they use the burn-in procedure described following Equation (595) to get their simulations going, whereas we use the stationary initial conditions advocated by Kay (1981b) and described in detail in Section 11.1 (see the discussion leading up to Equation (597a)). In the context of a Burg AR(p) estimate $\hat{S}(\cdot)$, their bootstrapping scheme consists of the following steps.

In the first step, given the Burg estimates $\hat{\phi}_{p,1}, \hat{\phi}_{p,2}, \dots, \hat{\phi}_{p,p}$ of the AR coefficients, we substitute these for $\phi_{p,1}, \phi_{p,2}, \dots, \phi_{p,p}$ in Equation (597b) with t set to p to get estimates $\hat{\phi}_{p-1,1}, \hat{\phi}_{p-1,2}, \dots, \hat{\phi}_{p-1,p-1}$ of the coefficients for the best linear predictor of order $p - 1$. We then use these coefficients in Equation (597b) with t now set to $p - 1$ to get estimated coefficients $\hat{\phi}_{p-2,1}, \hat{\phi}_{p-2,2}, \dots, \hat{\phi}_{p-2,p-2}$ for the best linear predictor of order $p - 2$. Continuing in this manner, we end up with Burg-based estimated coefficients $\hat{\phi}_{k,1}, \dots, \hat{\phi}_{k,k}$ for the best linear predictor of orders $k = p - 1, p - 2, \dots, 1$. We now have everything needed to compute the observed prediction errors $\overline{Z}_t(t)$, $t = 0, 1, \dots, N - 1$ as per Equation (618a). Given the Burg estimate $\hat{\sigma}_p^2$ of σ_p^2 , we can also compute the normalized observed prediction errors $\overline{Z}_t(t)$ as per Equation (618b). (As noted in Section 9.5, Burg's algorithm is recursive. A look at how its recursions work reveals that code for it could extract all the required prediction errors along with the $\hat{\phi}_{j,j}$ coefficients on the fly. If the code were to return these along with the $p + 1$ Burg parameter estimates, it would be an easy task to compute the normalized observed prediction errors $\overline{Z}_t(t)$.)

In the second step, given the Burg coefficient estimates and the normalized observed prediction errors $\{\overline{Z}_t(t) : t = 0, 1, \dots, N - 1\}$, we now consider Equation (597a), which tells us how to simulate an AR(p) time series given (i) the coefficients of the best linear predictors of orders unity up to p ; (ii) the square roots of σ_t^2 , $t = 0, 1, \dots, p$; and (iii) a random sample $\{Z_t\}$ of size N from a standard Gaussian distribution. In the bootstrapping scheme, for (i), we substitute the estimated coefficients we just formed in the first step; for (ii), we use square roots of the estimates $\hat{\sigma}_t^2$ defined in Equation (618b); and, most importantly, for (iii), we substitute a random sample of size N with replacement from the normalized observed prediction errors $\{\overline{Z}_t(t)\}$. Since we are using $\{X_t\}$ to denote the ocean wave data, let X'_t , $t = 0, 1, \dots, N - 1$, denote the simulated series we get from the right-hand sides of Equation (597a).

In the third step, we use $\{X'_t\}$ in exactly the same manner in which we used $\{X_t\}$ to

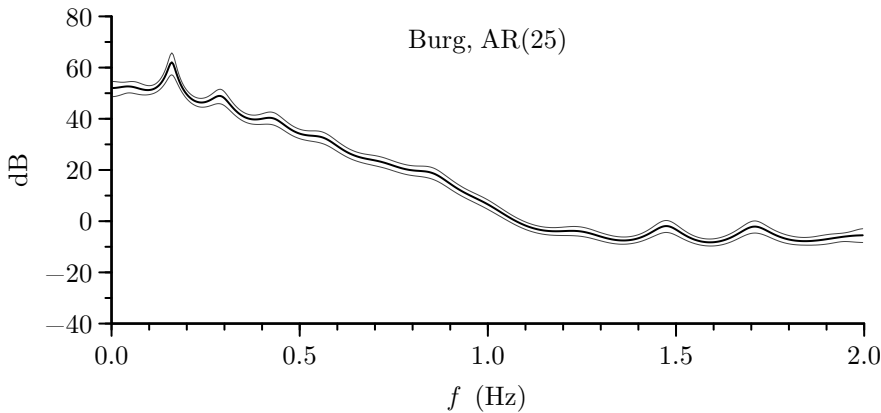


Figure 625a Burg AR(p) spectral estimate $\hat{S}(\cdot)$ of order $p = 25$ for ocean wave data (dark curve, reproduced from Figures 500(a) and (b)). The thin curves surrounding the dark curve show 95% CIs based upon the bootstrapping scheme of Swanepoel and Van Wyk (1986).

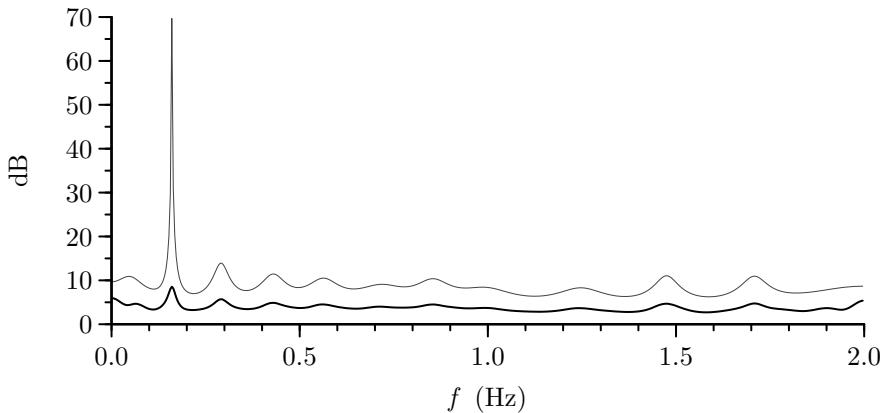


Figure 625b Widths of bootstrapped 95% CIs shown as thin curves in Figure 625a (dark curve), along with widths of no less than 95% CIs shown as thin curves in Figure 500(a) (light curve).

form an AR(25) Burg SDF estimate (in particular, we subtract off the sample mean of $\{X'_t\}$ as a preprocessing step). Denote this estimate by $\hat{S}_1(\cdot)$ (for comparison with the actual SDF estimate $\hat{S}(\cdot)$ for the ocean wave series, we compute $\hat{S}_1(\cdot)$ over all Fourier frequencies $f_k = k/(N \Delta_t)$ satisfying $0 < f_k < f_N$). We then generate a second simulated series in the same manner as the first, but using a different random sample with replacement from $\{\vec{Z}_t(t)\}$. We use this to create another AR(25) SDF estimate, namely, $\hat{S}_2(\cdot)$. We repeat this procedure over and over again, ending up with a large number of AR(25) SDF estimates $\hat{S}_n(\cdot)$, $n = 1, 2, \dots, N_R$, where here we set $N_R = 100,000$.

In the fourth and final step, we form a bootstrapped pointwise 95% CI for the unknown $S(f_k)$ by sorting the $\hat{S}_n(f_k)$ estimates – call these $\hat{S}_{(n)}(f_k)$ so that $\hat{S}_{(n)}(f_k) \leq \hat{S}_{(n+1)}(f_k)$. We then use $\hat{S}_{(\lfloor 0.975 N_R \rfloor)}(f_k)$ and $\hat{S}_{(\lfloor 0.025 N_R \rfloor)}(f_k)$ to define the upper and lower bounds of the 95% CI.

Figure 625a shows the bootstrapped 95% CIs for $S(f_k)$ versus f_k as two light curves above and below a dark curve depicting the Burg AR(25) estimate for the ocean wave series. The Burg estimate $\hat{S}(f_k)$ is trapped between the upper and lower CI at all Fourier frequencies

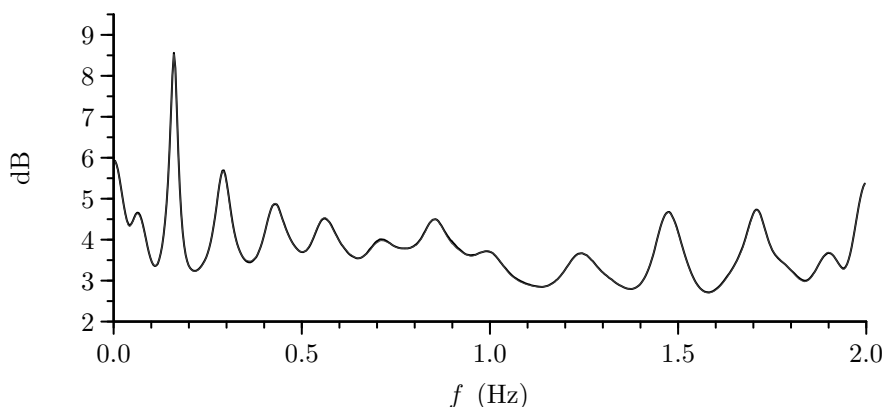


Figure 626 Widths of bootstrapped 95% CIs shown as thin curves in Figure 625a (dark curve, replicated from Figure 625b). These CIs make use of a random sample with replacement from the normalized observed prediction errors $\{\tilde{Z}_t(t)\}$. There is a light curve showing corresponding widths when using a random sample $\{Z_t\}$ from a standard Gaussian distribution, but this curve is virtually identical to the dark curve (the maximum absolute deviation between the two curves is 0.06 dB).

satisfying $0 < f_k < f_N$. A visual comparison of Figure 625a with 500(a) suggests that the bootstrapped 95% CIs are tighter than the no less than 95% CIs based on the large-sample analytic theory presented in Section 9.9. Figure 625b shows that this is indeed the case. The dark curve plots the widths of the bootstrapped CIs, while the light curves does the same for the analytic CIs. The widths of the analytic CIs are systematically larger, particularly at $f_{41} \doteq 0.16$ Hz. The median of the dB widths for the bootstrapped CIs is 3.7 dB, and it is 8.3 dB for the analytic CIs. In stark contrast to Figure 623b, with its close agreement between the bootstrapped CIs based on the Gaussian lag window estimate and ones based on large-sample theory, Figure 625b indicates poor agreement, possibly attributable to the “no less than 95%” nature of the analytic CIs (thus resulting in wider CIs than true 95% CIs).

There is a second way to do bootstrapping in the AR context. We do exactly as before, with the sole change being to use a random sample $\{Z_t\}$ from a standard Gaussian distribution rather than a random sample with replacement from the normalized observed prediction errors $\{\tilde{Z}_t(t)\}$. This second way is closer in spirit to the circulant embedding approach used with the Gaussian lag window estimate in that both use $\{Z_t\}$. Figure 626 shows that this second way yields CI widths that are remarkably similar to those of the first. One explanation for this close match is that $\{\tilde{Z}_t(t)\}$ is indistinguishable from deviates drawn from a standard Gaussian distribution, so it does not matter if we sample from $\{\tilde{Z}_t(t)\}$ or from $\{Z_t\}$.

In conclusion, simulated ocean wave series are useful not only for mimicking potentially important aspects of the original series, but also in analyzing the series. On the analysis side, simulated series help assess the assumption of Gaussianity and the variability of nonparametric and parametric SDF estimates. In the case of a nonparametric lag window SDF estimate, the assessment of variability via simulated series agrees remarkably well with that provided by the large-sample theory of Section 7.4. In the case of a parametric AR estimate, assessments from simulated series indicate much less variability than what the theory presented in Section 9.9 suggests. As noted in that section, this theory is inherently conservative. Simulated series verify this conservativeness and provide a second – arguably more realistic – take on the variability in a parametric AR estimate.

Ship Altitude Data

As an interesting contrast to the ocean wave data, let us consider the ship altitude time series

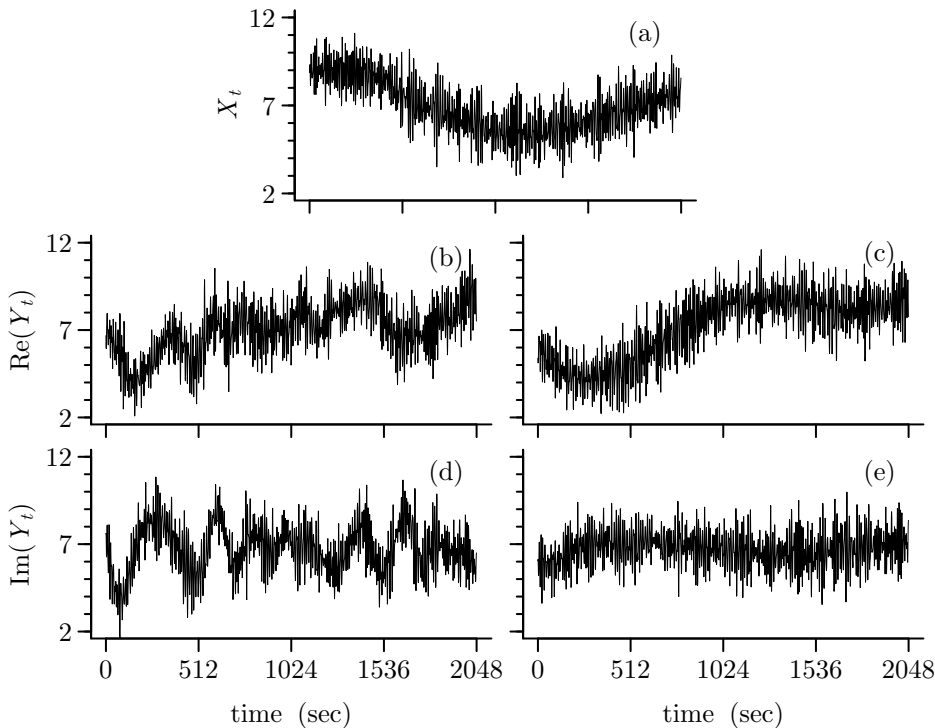


Figure 627 Ship altitude time series $\{X_t : t = 0, 1, \dots, N - 1\}$ (plot (a) – also shown in Figure 330) and four simulations of this series based upon circulant embedding (plots (b) to (e)). Here $N = 2048$, and the data have a sampling interval of $\Delta_t = 1$ sec. For plots (b) and (d), the weights S_k , $k = 0, 1, \dots, 2N - 1$, used to create the complex-valued series $\{Y_t : t = 0, 1, \dots, 2N - 1\}$ of Equation (602) are based on the $m = 80$ Gaussian lag window estimate $\hat{S}_m^{(LW)}(\cdot)$ shown as the dark curve in Figure 331(c); (b) and (d) show, respectively, the real and imaginary parts of Y_t , $t = 0, 1, \dots, N - 1$. Plots (c) and (e) are similar, but now use the Hanning-based direct spectral estimate $\hat{S}^{(D)}(\cdot)$ shown as the dark curve in Figure 331(b). The same sample of $4N$ deviates from a standard Gaussian distribution was used in creating the Y_t series from both $\hat{S}_m^{(LW)}(\cdot)$ and $\hat{S}^{(D)}(\cdot)$.

$\{X_t\}$ as a second example of simulating a time series using circulant embedding. This series has a sampling interval of $\Delta_t = 1$ sec and is shown in Figure 627(a), which is a replication of Figure 330. Figure 331 shows several nonparametric SDF estimates for this series. Leakage is a concern for some of these estimates. The most attractive estimate is the $m = 80$ Gaussian lag window estimate $\hat{S}_m^{(LW)}(\cdot)$ shown as the dark curve in Figure 331(c). This estimate is a smoothed version of the Hanning-based direct spectral estimate $\hat{S}^{(D)}(\cdot)$ shown in 331(b). The Hanning data taper is useful here because the periodogram shows evidence of bias due to leakage, and the Gaussian lag window is also beneficial because the resulting $\hat{S}_m^{(LW)}(\cdot)$ shows no evidence of bias due to smoothing window leakage. Under the assumption that the ship altitude series is a realization of a stationary process, the nonparametric SDF estimate $\hat{S}_m^{(LW)}(\cdot)$ is potentially a good basis for simulating time series whose statistical properties are in agreement with those of the actual series.

Plots (b) and (d) of Figure 627 shows two simulations of $\{X_t\}$ using circulant embedding in conjunction with the Gaussian lag window estimate $\hat{S}_m^{(LW)}(\cdot)$. The weights S_k are formulated in a manner similar to what we described for the preceding ocean wave example; however, here the ACVS estimate $\hat{s}_\tau^{(D)}$ of Equation (188b) uses the Hanning taper of Equation (189b), but, prior to forming the estimate, we again center the time series by subtracting off its sample mean, which here is $\bar{X} \doteq 6.943$. To facilitate comparison of the simulated

series with the actual time series, we have added \bar{X} to both simulated series prior to plotting them in Figures 627(b) and (d). This addition results in the sample mean for the series of length $N = 2048$ in plot (b) being 7.002; it is 6.733 in plot (d).

A visual comparison of the simulated series in plots (b) and (d) of Figure 627 with the ship altitude time series in plot (a) is somewhat disconcerting. While the high-frequency fluctuations seem comparable, the low-frequency ones are not. The actual series has a prominent low-frequency component spanning the entire time series, whereas the simulated series do not have such a prominent pattern. Is there an explanation for this mismatch? Suppose, for the sake of argument, that the true SDF has a narrow peak at frequency $1/2048$ Hz, which is associated with a period of 2048 sec (the time span of the entire time series). Consider the bandwidth of the Gaussian lag window estimate $\hat{S}_m^{(LW)}(\cdot)$, namely, $B_U \doteq 0.01$ Hz (calculated as per Equation (256a)). Due to this bandwidth, the peak would appear in $\hat{S}_m^{(LW)}(\cdot)$ as smeared out roughly over frequencies ranging from 0 to $1/2048 \text{ Hz} + B_U/2 \doteq 0.0055$ Hz. The uppermost frequency in this range has a period corresponding to 182 sec. Thus simulations based on $\hat{S}_m^{(LW)}(\cdot)$ would not exhibit a single component of a period 2048 sec, but rather a mishmash of components associated with periods of 182 sec and higher. Such a mishmash arguably would look like the low-frequency patterns in the simulated series in plots (b) and (d).

If the mishmash explanation is correct, creating simulations using an SDF estimate with a narrower bandwidth than that of $\hat{S}_m^{(LW)}(\cdot)$ might generate low-frequency patterns that are more in keeping with those of the ship altitude series. Now the Gaussian lag window estimate $\hat{S}_m^{(LW)}(\cdot)$ is a smoothed version of the Hanning-based direct spectral estimate $\hat{S}_m^{(LW)}(\cdot)$, which appears to be leakage free and has an associated bandwidth of $B_H \doteq 0.001$ Hz (calculated as per Equation (194)). The bandwidth B_H is a tenth of the size of B_U . The smearing of the peak would now go out to frequency $1/2048 + B_H/2 \doteq 0.0010$ Hz, which corresponds to a period of 1004 sec, i.e., about half the span of the entire time series. Plots (c) and (e) of Figure 627 shows two simulations of $\{X_t\}$ using circulant embedding in conjunction with the Hanning-based estimate $\hat{S}^{(D)}(\cdot)$ shown in Figure 331(b), with the weights S_k now being dictated by Equation (612a). The random sample of size $4N = 8192$ from a standard Gaussian distribution used to create these two simulated series is the same as was used to create the lag window-based simulations in plots (b) and (d) (reusing the random sample makes the simulated series such that any differences between them are solely due to their weighing schemes). The low-frequency patterns in the simulated series in plots (c) and (e) seem to match reasonably well with that of the ship altitude time series in plot (a). This result lends credence to the hypothesis that the reason why the series based on $\hat{S}_m^{(LW)}(\cdot)$ fail to replicate the ship altitude data at the very lowest frequencies is due to the bandwidth associated with $\hat{S}_m^{(LW)}(\cdot)$.

To conclude, while the Gaussian lag window estimate $\hat{S}_m^{(LW)}(\cdot)$ does a credible job for the most part, its bandwidth is potentially a detriment to reliable estimation of the SDF at the very lowest frequencies. Visual inspection of simulations using circulant embedding based on $\hat{S}_m^{(LW)}(\cdot)$ brings out this mismatch between the time series and the SDF estimate. Caution must be exercised in using simulated series based on $\hat{S}_m^{(LW)}(\cdot)$ if the purpose of the simulation includes matching the properties of the ship altitude series at the very lowest frequencies. Of course, at higher frequencies than these, there are regions of potential interest in the SDF. For example, there is a curious dip centered near 0.05 Hz that is evident in all the estimated SDFs in Figure 331. We could use simulations based on $\hat{S}_m^{(LW)}(\cdot)$ to investigate hypotheses about this dip.

Atomic Clock Data

In Sections 8.9 and 9.12 we looked at estimating the innovation variance for the atomic clock fractional frequency deviates $\{Y_t\}$ using, respectively, multitaper and AR approaches (see Figure 326(b) for a plot of $\{Y_t\}$ itself). The bottom row of Table 432 lists the multitaper-

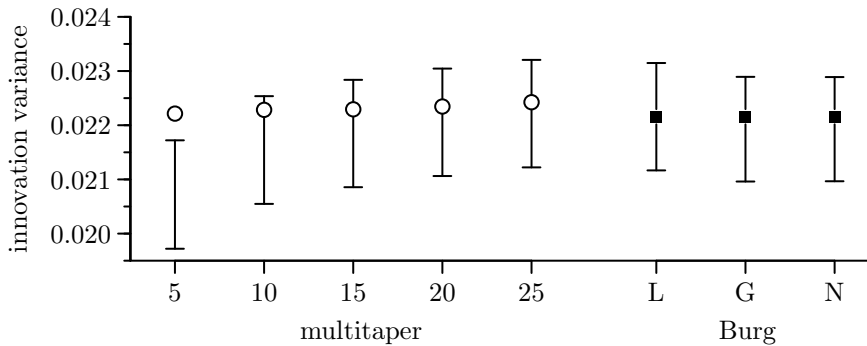


Figure 629 Multitaper-based estimates (circles) and Burg-based AR estimates (squares) of innovation variance for atomic clock fractional frequency deviates $\{Y_t\}$. The five displayed multitaper-based estimates differ in the number K of sinusoidal tapers used to create the estimate (5, 10, ..., 25). The three displayed AR estimates are identical – each is based on Burg’s algorithm with $p = 42$. The vertical lines (all but one interrupted by a circle or square) depict 95% CIs for the unknown theoretical innovation variance. The CIs associated with multitapering employ simulations based on circulant embedding. The first (“L”) of the three Burg-based CIs makes use of large-sample theory; the second (“G”), of simulated AR series driven by samples from a standard Gaussian distribution; and the third, also of simulated series, but now driven by normalized observed prediction errors.

based estimates $\hat{\sigma}_{(\text{MT})}^2$ using sinusoidal multitapering with $K = 5, 10, 15, 20$ and 25 tapers; these are also shown by circles in Figure 629. These five estimates agree well with one another (the ratio of the largest to the smallest is 1.01). The bottom row of Table 500 lists AR estimates based on the Yule–Walker method and Burg’s algorithm with p set to 42 in both cases. Both estimates agree well with the multitaper estimates; for simplicity, we focus here just on Burg. Each square in Figure 629 portrays the Burg estimate and has a common vertical displacement of $\hat{\sigma}_p^2 \doteq 0.02216$. Large-sample statistical theory suggests that, based on the Burg estimate, an approximate 95% CI for the unknown theoretical innovation variance is $\hat{\sigma}_p^2 \pm 1.96 \times \hat{\sigma}_p^2 \sqrt{(2/N)}$, where $N = 3999$ for the fractional frequency series. The resulting CI is $[0.02119, 0.02313]$, which is portrayed by the vertical line intersecting the left-most square in Figure 629.

A large-sample statistical theory has yet to be formulated that would allow construction of similar CIs for multitapering. In lieu of suitable theory, let us explore CIs based on simulated series with statistical properties in agreement with the multitaper SDF estimate used to form a particular $\hat{\sigma}_{(\text{MT})}^2$. In addition, we use simulated series to check the reasonableness of the CI based on the Burg estimate $\hat{\sigma}_p^2$ and large-sample theory.

Turning first to multitapering, let $\hat{S}^{(\text{MT})}(\cdot)$ denote the sinusoidal multitaper SDF estimate associated with one of the five estimates $\hat{\sigma}_{(\text{MT})}^2$ of the innovation variance shown as circles in Figure 629 (thus $\hat{S}^{(\text{MT})}(\cdot)$ employs either $K = 5, 10, 15, 20$ or 25 tapers). We can use circulant embedding to simulate series whose theoretical SDF is the same as $\hat{S}^{(\text{MT})}(\cdot)$. Since $\Delta_t = 1$ minute for $\{Y_t\}$, we can just ignore the sampling time and take the weights S_k needed by circulant embedding to be $\hat{S}^{(\text{MT})}(\tilde{f}_k)$, $0 \leq k \leq 2N - 1$, where $\tilde{f}_k = k/(2N)$. This is in agreement with Equation (613a) if we set $d_{k'} = 1/K$ and take \tilde{X}_t of Equation (611a) to be $Y_t - \bar{Y}$, i.e., the fractional frequency deviates centered about their sample mean. The weights, along with a random sample of size $4N = 15,996$ from a standard Gaussian distribution, go into the construction of the \mathcal{Y}_k series of Equation (601c). The Fourier transform of $\{\mathcal{Y}_k\}$ yields the complex-valued series $Y_t = Y_{\Re,t} + iY_{\Im,t}$ of Equation (602) (note carefully that, while Equation (602) uses Y_t to denote this complex-valued series, it is *not* the same as the real-valued fractional frequency deviates, which unfortunately are also denoted by Y_t). The series $Y_{\Re,t}$ and $Y_{\Im,t}$, $t = 0, 1, \dots, N - 1$, are two independent realizations of a zero mean

Gaussian process whose theoretical SDF is $\hat{S}^{(\text{MT})}(\cdot)$. After centering each series by subtracting off their respective sample means, we then compute multitaper SDF estimates using the same K sinusoidal tapers that went into creating $\hat{S}^{(\text{MT})}(\cdot)$. These two new multitaper SDF estimates are then fed into Equation (405b) to produce two new multitaper-based estimates of the innovation variance. We repeat this procedure a large number of times (each with a different random sample of size $4N$), ending up with N_{R} estimates of the innovation variance (we set $N_{\text{R}} = 100,000$). We then sort these estimates, index them by $1, 2, \dots, N_{\text{R}}$ and use the values indexed by 2,500 and 97,500 as lower and upper limits for a 95% CI for the unknown theoretical innovation variance. Figure 629 depicted these limits for the five settings of K . The heights of these CIs and the one based upon large-sample theory for the Burg estimate (surrounding the left-most square) are quite similar, differing at most by 4%; however, the lower and upper CI limits are markedly different, with the $K = 5$ multitaper-based CI not even trapping the corresponding estimate! The downward displacements of the CIs relative to their corresponding estimates for small settings of K flag these CIs as questionable.

Turning now to the AR case, we consider two schemes for creating a 95% CI for the unknown theoretical innovation variance using the $p = 42$ Burg-based parameter estimates $\bar{\phi}_{p,1}, \bar{\phi}_{p,2}, \dots, \bar{\phi}_{p,p}$ and $\bar{\sigma}_p^2$ for $\{Y_t\}$ after centering. For both schemes, we take the p coefficient estimates and recursively use Equation (597b) to obtain corresponding estimates $\bar{\phi}_{j,k}$ for the coefficients of the best linear predictors of orders $j = 1, 2, \dots, p-1$. We then substitute $\bar{\phi}_{j,j}$ and $\bar{\sigma}_p^2$ for $\hat{\phi}_{j,j}$ and $\hat{\sigma}_p^2$ in Equation (618b) to obtain an estimate $\bar{\sigma}_t^2$ for σ_t^2 , $0 \leq t \leq p-1$. The first scheme is based upon Equation (597a) with the unknown $\phi_{j,k}$ and σ_t^2 terms replaced by estimates $\bar{\phi}_{j,k}$ and $\bar{\sigma}_t^2$. Using these substitutions, we use a sample $\{Z_t\}$ of length N from a standard Gaussian distribution to create a simulated series $\{X_t\}$ to be regarded as a realization of a Gaussian AR(42) process with theoretical parameters identical to the Burg parameter estimates. We then take $\{X_t\}$, center it using its sample mean and feed the centered series into Burg's algorithm to get a new estimate of the innovation variance σ_p^2 . We repeat this procedure N_{R} times (each time using a different random sample of size N), ending up with $N_{\text{R}} = 100,000$ Burg-based estimates of the innovation variance. We then sort these estimates and use those indexed by 2,500 and 97,500 as the lower and upper limits of a 95% CI for the unknown theoretical σ_p^2 . Figure 629 shows these limits emanating from the middle square. The height of this CI is similar to the one to its immediate left, which was obtained from large-sample theory; however, the latter CI is symmetric about $\bar{\sigma}_p^2$, but the former is not.

The second AR scheme for creating a 95% CI is *exactly* the same as the first, with one exception: in the spirit of bootstrapping, instead of using a random sample $\{Z_t\}$ from a standard Gaussian distribution in Equation (597a), we take instead a random sample with replacement of size N from the N normalized observed prediction errors $\{\bar{Z}_t(t)\}$ formed as per Equation (618b) (we interpret $\hat{\phi}_{j,j}$ and $\hat{\sigma}_p^2$ as the Burg estimates $\bar{\phi}_{j,j}$ and $\bar{\sigma}_p^2$ obtained from $\{Y_t\}$ after centering). The resulting CI is the right-most one depicted in Figure 629. This CI is virtually identical to the one to its immediate left, which comes from the first AR scheme. The fact that the two schemes yield such similar results is presumably attributable to $\{\bar{Z}_t(t)\}$ being a good approximation to a random sample from a standard Gaussian distribution.

Thus, as noted in Section 9.12, the nonparametric multitaper-based estimates of the innovation variance and the parametric Burg-based AR estimates agree reasonably well. In addition, simulation-based CIs for the unknown theoretical innovation variance have remarkably similar heights in the nonparametric and parametric cases (and all agree well with the height of the CI determined by large-sample theory for the Burg-based estimate of σ_p^2); however, while the CIs themselves are in reasonable agreement for the $K = 20$ and 25 multitaper-based cases and the three Burg-based cases, the $K = 5$ and 10 multitaper-based CIs are markedly out of sync and intuitively unappealing.

Comments and Extensions to Section 11.6

[1] In the examples presented in this section, we looked at two different methods for simulating time series that fit in with the notion of bootstrapping: a parametric method based on the AR model (Swanepoel and Van Wyk, 1986) and a nonparametric method based on circulant embedding (Percival and Constantine, 2006). While these methods allow for bootstrapping certain statistics associated with time series, they are certainly not the only methods that do so. Comprehensive reviews of the literature on bootstrapping for time series are Politis (2003) and Kreiss and Paparoditis (2011). The books by Davison and Hinkley (1997) and Lahiri (2003b) discuss both parametric and nonparametric methods. A critique of some of the nonparametric methods discussed in these references is an over-reliance on asymptotic properties of the periodogram, particularly regarding unbiasedness and independence across the grid of Fourier frequencies (as we have emphasized throughout this book, these are dicey assumptions in some practical applications). Use of circulant embedding with nonparametric SDF estimates other than the periodogram can potentially overcome the problems associated with the periodogram. Circulant embedding worked credibly in some – but not all – of the examples we presented (see the discussion surrounding Figure 629), illustrating that more work is needed to establish its limitations.

11.7 Comments on Simulation of Non-Gaussian Time Series

Simulation of a non-Gaussian time series having a specified ACVS or SDF structure is generally considerably more involved than the Gaussian case discussed extensively in this chapter. One approach for ARMA processes (Davies et al., 1980) is to generate an IID noise sequence having a distribution such that, when it is subjected to the filter required to obtain the specified ARMA form, the marginal distribution of the filter output has the desired first four moments. An alternative is to start with IID Gaussian noise and pass it through both a linear filter and a nonlinear transform; the correct design of both these components yields the desired second-order characteristics (ACVS, SDF) and marginal distribution. Note that in the first approach the aim is to produce a *constrained* marginal distribution having specified first four moments, and in the second case, a *full* marginal distribution, but in both cases only first-order PDF modeling is attempted.

For the first approach, the desire is to simulate a zero mean non-Gaussian, causal, ARMA process. Let $\{\epsilon_t\}$ denotes a zero mean IID sequence with variance σ_ϵ^2 (a stronger assumption than white noise, but IID implies that the RVs are uncorrelated). We can generate such a sequence in practice by simply sampling independent RVs from a suitable zero mean distribution and rescaling if necessary. Then for the causal ARMA process there exists a unique set of weights $\{\psi_j\}$ such that

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \quad \text{with } \psi_0 \stackrel{\text{def}}{=} 1$$

(this is Equation (600c)). The IID sequence is passed through the linear time-invariant (LTI) filter $\{\psi_j\}$ to produce the zero mean linear process $\{X_t\}$. Since $\{\epsilon_t\}$ is strictly stationary, the output $\{X_t\}$ of the LTI filter is also strictly stationary (Papoulis and Pillai, 2002, section 9.2). In particular, each RV ϵ_t in the sequence $\{\epsilon_t\}$ has the same marginal distribution, and likewise each X_t in $\{X_t\}$ has the same marginal distribution, which is in general different from that of ϵ_t and is intended to be non-Gaussian.

Now $\{\epsilon_t\}$ is IID with variance σ_ϵ^2 , so $\sigma^2 \stackrel{\text{def}}{=} \text{var}\{X_t\} = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j^2$. For the zero mean RV X_t , the moment coefficient of skewness $\sqrt{\delta_1(X_t)}$ and moment coefficient of kurtosis $\delta_2(X_t)$ are given by

$$\sqrt{\delta_1(X_t)} \stackrel{\text{def}}{=} \frac{E\{X_t^3\}}{(E\{X_t^2\})^{3/2}} = \frac{E\{X_t^3\}}{\sigma^3} \quad \text{and} \quad \delta_2(X_t) \stackrel{\text{def}}{=} \frac{E\{X_t^4\}}{(E\{X_t^2\})^2} = \frac{E\{X_t^4\}}{\sigma^4},$$

which leads to

$$\sqrt{\delta_1(X_t)} = \frac{\sum_{j=0}^{\infty} \psi_j^3}{(\sum_{j=0}^{\infty} \psi_j^2)^{3/2}} \sqrt{\delta_1(\epsilon_t)} \quad (632a)$$

and

$$\delta_2(X_t) = \frac{\sum_{j=0}^{\infty} \psi_j^4}{(\sum_{j=0}^{\infty} \psi_j^2)^2} \delta_2(\epsilon_t) + 6 \frac{\sum_{j=0}^{\infty} \sum_{i>j} \psi_i^2 \psi_j^2}{(\sum_{j=0}^{\infty} \psi_j^2)^2} \quad (632b)$$

(Davies et al., 1980). Since X_t has zero mean, its first four cumulants are given by $\kappa_1(X_t) = 0$, $\kappa_2(X_t) = E\{X_t^2\} = \sigma^2$, $\kappa_3(X_t) = E\{X_t^3\}$ and $\kappa_4(X_t) = E\{X_t^4\} - 3\sigma^4$.

▷ **Exercise [632a]** Show that

$$\delta_2(X_t) - 3 = \frac{\sum_{j=0}^{\infty} \psi_j^4}{(\sum_{j=0}^{\infty} \psi_j^2)^2} [\delta_2(\epsilon_t) - 3] \text{ and also } \frac{\kappa_4(X_t)}{\sigma^4} = \frac{\sum_{j=0}^{\infty} \psi_j^4}{(\sum_{j=0}^{\infty} \psi_j^2)^2} \frac{\kappa_4(\epsilon_t)}{\sigma_{\epsilon}^4}. \quad (632c) \triangleleft$$

Hence the scaled fourth-order cumulant of X_t is equal to the scaled fourth-order cumulant of ϵ_t multiplied by the factor $\sum_{j=0}^{\infty} \psi_j^4 / (\sum_{j=0}^{\infty} \psi_j^2)^2$, a quantity called the kurtosis norm of the filter $\{\psi_j\}$ in Longbottom et al. (1988). Here are some notes:

- [1] If either $\sqrt{\delta_1(\epsilon_t)} = 0$ or $\sum_{j=0}^{\infty} \psi_j^3 / (\sum_{j=0}^{\infty} \psi_j^2)^{3/2} = 0$, we see from Equation (632a) that $\sqrt{\delta_1(X_t)} = 0$. So zero-skewness noise gives rise to a zero-skewness process, but this is not the only way that the latter can occur.
- [2] Equation (632c) shows that $\delta_2(X_t) = 3$ if and only if $\delta_2(\epsilon_t) = 3$. So a kurtosis of 3 for ϵ_t is a necessary and sufficient condition for a kurtosis of 3 for X_t .
- [3] Since $(\sum_{j=0}^{\infty} \psi_j^2)^2 \geq \sum_{j=0}^{\infty} \psi_j^4$, the kurtosis norm of the filter is bounded by unity, and so the scaled fourth cumulant of X_t never exceeds the scaled fourth cumulant of ϵ_t (the solution to Exercise [632a] makes this inequality obvious, and shows the inequality is strict unless all $\psi_j = 0$ for $j > 0$, i.e., unless $\{X_t\}$ is IID noise).

Properties [1] and [2] are consistent with filtered independent Gaussian noise being itself Gaussian. Property [3] tells us that, in the sense of the scaled fourth cumulant, the distribution of the filtered IID noise X_t is closer to Gaussian than that of the IID noise itself (recall that $\kappa_4(X_t)$ would be zero if X_t were Gaussian).

As an example, in the case of an ARMA(1,1) process, $X_t = \phi X_{t-1} + \epsilon_t + \vartheta \epsilon_{t-1}$, the forms of the sums in Equations (632a) and (632c) can be readily evaluated analytically, leading to the following results.

▷ **Exercise [632b]** Show that, for an ARMA(1,1) process,

$$\text{var}\{X_t\} = \sigma^2 = \sigma_{\epsilon}^2 [1 - \phi^2 + (\phi + \vartheta)^2] / (1 - \phi^2), \quad (632d)$$

while Equation (632a) becomes

$$\sqrt{\delta_1(X_t)} = \frac{[1 - \phi^3 + (\phi + \vartheta)^3](1 - \phi^2)^{3/2}}{[1 - \phi^2 + (\phi + \vartheta)^2]^{3/2}(1 - \phi^3)} \sqrt{\delta_1(\epsilon_t)} \quad (632e)$$

and Equation (632c) leads to

$$\delta_2(X_t) - 3 = \frac{[1 - \phi^4 + (\phi + \vartheta)^4](1 - \phi^2)}{[1 - \phi^2 + (\phi + \vartheta)^2]^2(1 + \phi^2)} [\delta_2(\epsilon_t) - 3]. \quad (632f) \triangleleft$$

If we wish to simulate a sequence $\{X_t\}$ having mean zero, variance σ^2 , skewness $\sqrt{\delta_1(X_t)}$ and kurtosis $\delta_2(X_t)$ and having a causal ARMA(1,1) structure, then we need to generate IID noise $\{\epsilon_t\}$ with mean zero, variance satisfying Equation (632d) and skewness and kurtosis satisfying Equations (632e) and (632f), respectively. This can be accomplished by generating an IID sequence from a distribution with these specified moment constraints. The Johnson system of distributions covers many combinations of skewness and kurtosis; see, e.g., Figure 2 of Johnson (1949). Hill (1976) and Hill et al. (1976) provide FORTRAN routines to calculate the Johnson parameters corresponding to a required mean, variance, skewness and kurtosis, which can be used to simulate Johnson RVs from Gaussian RVs (translations are also available in R).

Exercise [11.14] establishes that, in the ARMA(1,1) case, $|\sqrt{\delta_1(X_t)}| < |\sqrt{\delta_1(\epsilon_t)}|$ under the assumption $\phi + \vartheta \neq 0$; i.e., the size of the skewness of X_t is less than the size of the skewness of ϵ_t (when $\phi + \vartheta = 0$, the ARMA(1,1) process degenerates into IID noise; we then have $X_t = \epsilon_t$, and Equation (632e) says $|\sqrt{\delta_1(X_t)}| = |\sqrt{\delta_1(\epsilon_t)}|$). Thus, since the skewness of the Gaussian distribution is zero, if we only consider skewness, the distribution of the filtered IID noise X_t is closer to Gaussian than that of the IID noise itself (we reached a similar conclusion when we considered just the scaled fourth cumulant). The relationship between the PDFs for X_t and ϵ_t in cases other than ARMA(1,1) has been studied theoretically (Mallows, 1967; Granger, 1979).

Walden (1993) uses this general approach to synthesize sequences of primary reflection coefficients in the layered-earth model. Drilled wells at three locations enabled statistical model structures to be derived: ARMA(1,1) models with marginal distributions for $\{X_t\}$ having zero mean, zero skewness and kurtosis values of either 3.8, 4.7 or 8.1. Instead of the Johnson class, he uses the generalized Gaussian (GG) distribution as a model for the IID noise. The zero mean and symmetric GG has a PDF for ϵ_t given by

$$f(\epsilon) = \frac{\alpha A(\alpha, \sigma_\epsilon)}{2\Gamma(1/\alpha)} \exp(-[A(\alpha, \sigma_\epsilon)|\epsilon|]^\alpha), \quad |\epsilon| < \infty,$$

with $A(\alpha, \sigma_\epsilon) \stackrel{\text{def}}{=} \{\Gamma(3/\alpha)/[\sigma_\epsilon^2\Gamma(1/\alpha)]\}^{1/2}$, where α is the shape parameter, and $\Gamma(\cdot)$ is the gamma function. The shape parameter determines the peakedness at the origin and long-tailedness of the distribution relative to the Gaussian. The Laplace (double-sided exponential) distribution corresponds to $\alpha = 1$; the Gaussian, to $\alpha = 2$; and the uniform distribution is obtained as $\alpha \rightarrow \infty$. The zero mean and symmetric GG has

$$E\{\epsilon_t^{2n}\} = [\Gamma^{n-1}(1/\alpha)\Gamma([2n+1]/\alpha)\sigma_\epsilon^{2n}]/\Gamma^n(3/\alpha),$$

so that the value of α that gives the required kurtosis can be found. Simulation of GG variates can be accomplished using R code such as `pgnorm`. In Walden (1993) values of the four parameters input to the simulation, namely, variance and kurtosis, and ARMA parameters ϕ and ϑ , agree well with the corresponding estimated values from the simulated sequences.

We now turn to an alternative approach for the simulation of a non-Gaussian time series having a specified ACVS or SDF structure. This methodology is discussed in Gujar and Kavanagh (1968), Liu and Munson (1982) and Sondhi (1983). The approach can be illustrated thus:

$$\{\epsilon_t\} \longrightarrow \boxed{G(\cdot)} \longrightarrow \{X_t\} \longrightarrow \boxed{\mathcal{Z}(\cdot)} \longrightarrow \{Y_t\} \quad (633)$$

In the above $\{\epsilon_t\}$ is now taken to be a zero mean, unit variance, Gaussian IID sequence. The Gaussian IID sequence is passed through a LTI filter with transfer function $G(\cdot)$ to produce a linear Gaussian process $\{X_t\}$ with zero mean and ACVS $\{s_{X,\tau}\}$. The filter coefficients are

normalized so that the output has a unit variance, i.e., $s_{X,0} = 1$. The sequence $\{X_t\}$ is strictly stationary with mean zero and variance unity and is jointly Gaussian. This linear process is then subjected to a zero-memory (instantaneous) nonlinearity (ZMNL) denoted $\mathcal{Z}(\cdot)$ which is chosen so that the output process $\{\mathcal{Z}(X_t)\} \stackrel{\text{def}}{=} \{Y_t\}$ has a desired ACVS and marginal *non-Gaussian* PDF. Since $\{X_t\}$ is strictly stationary, $\{Y_t\}$, the output of the ZMNL, is itself strictly stationary (Papoulis and Pillai, 2002, section 9.2). Each RV X_t in the process $\{X_t\}$ has the same Gaussian marginal distribution, and each Y_t in $\{Y_t\}$ has the same *non-Gaussian* marginal distribution. Denote the Gaussian distribution function as usual by $\Phi(\cdot)$, and the distribution function of Y_t by $F_Y(\cdot)$. By the probability integral transform, the RV U resulting from transformation of the RVs X_t and Y_t via $U = \Phi(X_t)$ and $U = F_Y(Y_t)$ is uniformly distributed on $[0, 1]$. Then

$$\mathcal{Z}(X_t) \stackrel{\text{d}}{=} Y_t \stackrel{\text{d}}{=} F_Y^{-1}(U) \stackrel{\text{d}}{=} F_Y^{-1}(\Phi(X_t)).$$

This result shows how the form of the ZMNL is determined by the choice of $F_Y(\cdot)$. As an example suppose it is desired that the output $\{Y_t\}$ should have a marginal uniform distribution with mean zero and variance of unity, with PDF $f(y) = 1/(2\sqrt{3})$, $|y| \leq \sqrt{3}$. Then

$$F_Y(y) = \int_{-\sqrt{3}}^y [1/(2\sqrt{3})] dz = \frac{y + \sqrt{3}}{2\sqrt{3}},$$

so $U \stackrel{\text{d}}{=} (Y_t + \sqrt{3})/(2\sqrt{3})$ and therefore

$$\mathcal{Z}(X_t) \stackrel{\text{d}}{=} Y_t \stackrel{\text{d}}{=} [2U - 1]\sqrt{3} \stackrel{\text{d}}{=} [2\Phi(X_t) - 1]\sqrt{3}. \quad (634a)$$

Having derived the form of the nonlinearity, the next step is to find the relationship between $s_{X,\tau}$ and $s_{Y,\tau}$. Since $X_{t+\tau}$ and X_t both have mean zero, $s_{X,\tau} = E\{X_{t+\tau}X_t\}$, and since both RVs also have a variance of unity, $s_{X,\tau} = \rho_{X,\tau}$; i.e., the ACVS and ACS are identical for $\{X_t\}$. The bivariate standard Gaussian PDF for $(X_{t+\tau}, X_t)$ is

$$f(v, w; \rho_{X,\tau}) = \frac{1}{2\pi(1 - \rho_{X,\tau}^2)^{1/2}} \exp\left(-\frac{v^2 + w^2 - 2\rho_{X,\tau}vw}{2(1 - \rho_{X,\tau}^2)}\right).$$

If we choose to standardize $\{Y_t\}$ such that $s_{Y,0} = 1$ so that the ACVS and ACS are identical for $\{Y_t\}$ also, then $s_{Y,\tau} = \rho_{Y,\tau}$ and

$$\rho_{Y,\tau} \stackrel{\text{def}}{=} E\{Y_{t+\tau}Y_t\} - E\{Y_{t+\tau}\}E\{Y_t\} = E\{\mathcal{Z}(X_{t+\tau})\mathcal{Z}(X_t)\} - E\{\mathcal{Z}(X_{t+\tau})\}E\{\mathcal{Z}(X_t)\}.$$

Sondhi (1983) gives three examples where $\rho_{Y,\tau}$ can be found analytically. For Y_t having the uniform distribution with mean zero and variance of unity, $\mathcal{Z}(X_t)$ is given by Equation (634a), and Sondhi obtains

$$\rho_{Y,\tau} = (6/\pi) \sin^{-1}(\rho_{X,\tau}/2). \quad (634b)$$

A general approach for finding $\rho_{Y,\tau}$, developed in Liu and Munson (1982) and Sondhi (1983), makes use of Hermite polynomials. First,

$$E\{\mathcal{Z}(X_t)\} = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \mathcal{Z}(v)e^{-v^2/2} dv.$$

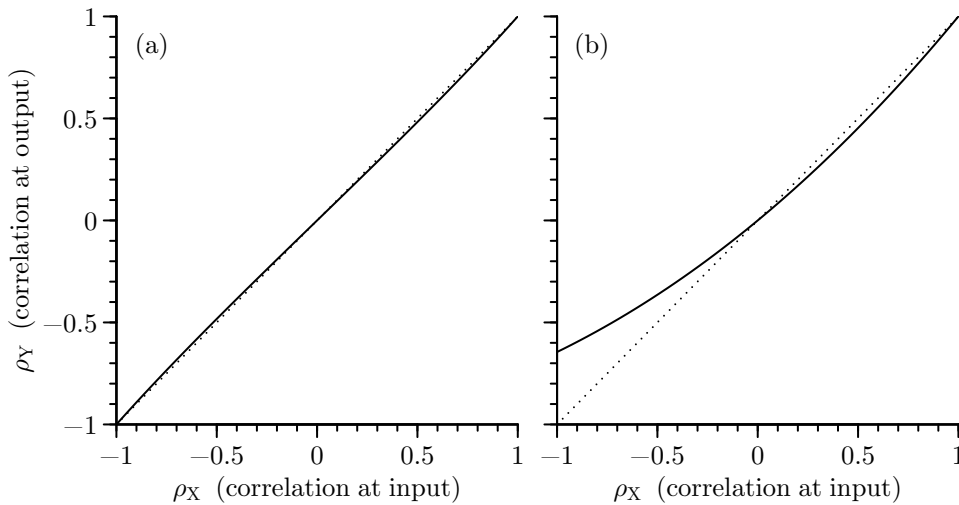


Figure 635 The relationship between the autocorrelation $\rho_{Y,\tau}$ at the output of the ZMNL and $\rho_{X,\tau}$ at the input. The dotted line marks equality. For (a) the distribution for Y_t is the uniform with mean zero and variance unity. The exact mapping given by Equation (634b) is shown by a dash-dot line, and that given by Equation (635d), using coefficients $\{d_k\}$ up to $k = 7$, is shown by a solid line; the two are indistinguishable. For (b) the distribution for Y_t is the exponential with variance unity. The mapping given by Equation (635d), using coefficients $\{d_k\}$ up to $k = 7$, is again shown by a solid line. Since in (a) the distribution is symmetric for Y_t , the relationship shows the expected odd symmetry. In (b) the distribution for Y_t is asymmetric, and the relationship does not show any recognizable symmetries.

Then $\mathcal{Z}(v)$ is expanded in terms of Hermite polynomials $\text{He}_k(v)$ with coefficients $\{d_k\}$,

$$\mathcal{Z}(v) = \sum_{k=0}^{\infty} d_k \frac{\text{He}_k(v)}{\sqrt{(k!)}}. \quad (635a)$$

The orthogonality properties of Hermite polynomials say that $E\{\mathcal{Z}(X_t)\} = d_0$, leading to

$$E\{\mathcal{Z}(X_{t+\tau})\}E\{\mathcal{Z}(X_t)\} = d_0^2.$$

Next,

$$E\{\mathcal{Z}(X_{t+\tau})\mathcal{Z}(X_t)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{Z}(v)\mathcal{Z}(w)f(v, w; \rho_{X,\tau}) \, dv \, dw. \quad (635b)$$

But the bivariate standard Gaussian PDF can also be expanded in terms of Hermite polynomials using Mehler's expansion (Barrett and Lampard, 1955):

$$f(v, w; \rho_{X,\tau}) = \frac{1}{2\pi} e^{-(v^2+w^2)/2} \sum_{k=0}^{\infty} \rho_{X,\tau} \frac{\text{He}_k(v)\text{He}_k(w)}{k!}. \quad (635c)$$

Equations (635a), (635b) and (635c) give $E\{\mathcal{Z}(X_{t+\tau})\mathcal{Z}(X_t)\} = \sum_{k=0}^{\infty} d_k^2 \rho_{X,\tau}^k$ (Wise et al., 1977). Finally,

$$\rho_{Y,\tau} = \sum_{k=0}^{\infty} d_k^2 \rho_{X,\tau}^k - d_0^2 = \sum_{k=1}^{\infty} d_k^2 \rho_{X,\tau}^k, \quad (635d)$$

and by setting $\tau = 0$ we see that $\sum_{k=1}^{\infty} d_k^2 = 1$.

Hence, given $\{\rho_{Y,\tau}\}$, Equation (635d) can be used to solve for $\{\rho_{X,\tau}\}$, and then $\{\epsilon_t\}$ can be correctly colored to produce $\{X_t\}$. Both Liu and Munson (1982) and Sondhi (1983) observe that, if the PDF of Y_t is symmetric, $\mathcal{Z}(\cdot)$ will be an odd function of its argument, and the even coefficients in $\{d_k\}$ vanish, making $\rho_{Y,\tau}$ an odd function of $\rho_{X,\tau}$; the relationship between $\rho_{Y,\tau}$ and $\rho_{X,\tau}$ is then invertible. Returning to the example of Y_t having the uniform distribution with mean zero and variance of unity, the invertibility can also be seen in Equation (634b). For this example, Liu and Munson (1982) obtain the following first few nonzero squared Hermite coefficients:

$$d_1^2 \doteq 0.9550627; \quad d_3^2 \doteq 0.0397943; \quad d_5^2 \doteq 0.0044768; \quad d_7^2 \doteq 0.0006662.$$

Using these values, we can compare the exact mapping between $\rho_{X,\tau}$ and $\rho_{Y,\tau}$ given by Equation (634b), with that given by Equation (635d), using coefficients $\{d_k\}$ up to $k = 7$. Figure 635(a) shows that there is no discernible difference between these two mappings. We also see that the choice to obtain a uniform distribution for the output of the ZMNL causes only a moderate distortion between $\rho_{X,\tau}$ and $\rho_{Y,\tau}$. Table 1 of Liu and Munson (1982) gives sets of (squared) Hermite coefficients for other distributions for Y_t . Taking those for the exponential distribution with variance unity, a highly asymmetric distribution, the relationship between $\rho_{X,\tau}$ and $\rho_{Y,\tau}$ can again be found from Equation (635d) and is given in Figure 635(b), which indicates a more substantial distortion between $\rho_{X,\tau}$ and $\rho_{Y,\tau}$. Our use of the term “distortion” is in line with Johnson (1994) calling this approach the *correlation distortion method*.

Liu and Munson (1982) discuss the condition for Equation (635d) to be solved for $\rho_{X,\tau}$ for arbitrary distributions for Y_t . They also point out that under some circumstances the derived $\{\rho_{X,\tau}\}$ may not be positive semidefinite as required (see Equation (28b)).

Suppose a solution to Equation (635d) has been found. Express it as $\hat{\rho}_{X,\tau}(\rho_{Y,\tau})$. (For the example of Y_t having the uniform distribution with mean zero and variance of unity, Equation (634b) gives the exact result $\hat{\rho}_{X,\tau}(\rho_{Y,\tau}) = 2 \sin(\pi \rho_{Y,\tau}/6)$.) The sequence $\{\hat{\rho}_{X,\tau}(\rho_{Y,\tau})\}$ is symmetric and its Fourier transform $\hat{S}_X(f)$, say, is real. For many choices of $\{\rho_{Y,\tau}\}$, it happens that $\hat{S}_X(f) \geq 0$ for all f , so that the required filter $\{g_u\}$ can be found by finding a transfer function $G(\cdot)$ such that $|G(f)|^2 = \hat{S}_X(f)$ (the procedure of finding a suitable “square root” for $\hat{S}_X(\cdot)$ is known as spectral factorization). If $\hat{S}_X(f) < 0$ at some frequency f , Sondhi (1983) suggests proceeding with $(\hat{S}_X(f))_+$ instead; practical algorithmic steps are provided. Liu and Munson (1982, p. 976) show that such an approach is often nearly optimal in minimizing distortion between the realized and desired output sequence $\{\rho_{Y,\tau}\}$.

Serroukh et al. (2000) use the simulation scheme in Equation (633) with three different non-Gaussian distributions for Y_t and a variety of ACS sequences $\{\rho_{Y,\tau}\}$ to test the theory associated with the wavelet variance estimator. Good agreement between simulation results and the expected theory was obtained.

One drawback with the scheme in Equation (633) is that the marginal PDF and ACS/SDF are linked together, so that changing one requires a change to the other. A novel method for both representing and generating realizations of a stationary non-Gaussian process is described by Kay (2010). Here specifications of the PDF and ACS/SDF are uncoupled, allowing independent specification of the first-order PDF of the random process and the autocorrelation structure; however, the PDF must be symmetric and infinitely divisible (this means that the desired PDF should correspond to that of a sum of an arbitrary number of IID random variables). By way of contrast, the scheme in Equation (633), when it can be implemented, has no such restrictions on the marginal PDF. Finally, Picinbono (2010) shows that, by using a structure with appropriate random coefficients, it is possible to construct time series with ARMA second-order properties, while the marginal PDF is symmetric, but otherwise unconstrained, and can be specified in advance.

11.8 Summary of Simulation of Time Series

Suppose we have either (i) a theoretical stationary process $\{X_t\}$ that is specified to some degree or (ii) a nonparametric or parametric estimate $\hat{S}(\cdot)$ of the spectral density function (SDF) based upon a realization of a portion X_0, X_1, \dots, X_{N-1} of a stationary process whose theoretical SDF is unknown. Suppose that, in case (i), we are interested in computer-generated artificial series (simulations) that can be regarded as realizations from $\{X_t\}$ and, in case (ii), in simulations whose statistical properties are in agreement with those of the estimate $\hat{S}(\cdot)$. There are a number of ways in which we can accomplish our goals, but these depend on exactly what we have to work with.

Turning first to case (i), suppose that $\{X_t\}$ is a zero mean autoregressive moving average (ARMA) process of order (p, q) (see Equation (35b)). We presume that its $p + q + 1$ parameters $\phi_{p,j}$, $\theta_{q,j}$ and σ_ϵ^2 are known so that $\{X_t\}$ is fully specified. If $\{X_t\}$ is in fact just white noise ($p = 0, q = 0$), it is an easy matter to generate simulations using computer routines purported to give random samples from a particular distribution (Gaussian is a popular choice, but white noise can be simulated easily with certain non-Gaussian distributions, e.g., uniform or exponential). If $\{X_t\}$ is a moving average (MA) process ($p = 0, q \geq 1$), the fact that it is a finite linear combination of white noise (not necessarily Gaussian) points to easy methods for generating simulations (see the discussion following Equation (594)). If $\{X_t\}$ is an autoregressive (AR) process ($p \geq 1, q = 0$), we face a challenge. By definition X_t involves a linear combination of X_{t-1}, \dots, X_{t-p} , so simulation of X_t would seem to require a simulation of X_{t-p} , which in turn needs a simulation of X_{t-2p} and so forth ad infinitum. This startup problem can be elegantly circumvented by generating stationary initial conditions (Kay, 1981b; see the discussion surrounding Equation (597a)). Using these conditions, we can simulate an AR series of length N based upon a random sample of size N , typically from a standard Gaussian distribution. If $\{X_t\}$ is an ARMA process with $p \geq 1$ and $q \geq 1$, we can merge the techniques for simulating AR and MA processes to come up with simulations of $\{X_t\}$ (see the discussion following Exercise [598]). If, on the other hand, the process of interest is a fully specified harmonic process, its formulations in terms of either random amplitudes (Equation (598a)) or random phases (Equation (598b)) suggest easy pathways to generate simulations.

Consider now another formulation of case (i), in which $\{X_t\}$ is specified either in terms of its autocovariance sequence (ACVS) $\{s_\tau\}$ or its SDF $S(\cdot)$. If s_0, s_1, \dots, s_N are known, we can use circulant embedding to simulate X_0, X_1, \dots, X_{N-1} subject to a set of weights S_k being nonnegative. The weights are obtained from the Fourier transform of a circularized version of the $N + 1$ known elements of the ACVS (see Equation (601a)). Unfortunately there are some stationary processes and some sample sizes N for which some of the weights are negative, which formally renders circulant embedding inapplicable (however, setting the offending weights to zero permits simulation of time series whose statistical properties are approximately consistent with the portion of the ACVS we know, with the degree of approximation being quantifiable). When the weights are nonnegative, circulant embedding is a computationally efficient method for generating exact simulations (this method is associated with a harmonic process, and here “exact” means that the ACVS for the harmonic process agrees perfectly with s_τ at the known lags). Its efficiency is due to its use of a Fourier transform involving the weights and a random sample of size $2N$, typically from a standard Gaussian distribution. (If circulant embedding cannot provide exact simulations because of negative weights, an alternative approach involves a Cholesky decomposition of a covariance matrix containing the known ACVS values, but this approach is computationally unappealing and is sometimes subject to numerical instabilities.)

If, on the other hand, $\{X_t\}$ is specified by its SDF $S(\cdot)$ rather than its ACVS, the Gaussian spectral synthesis method (GSSM) is worth exploring. Circulant embedding and GSSM use weights and random samples in exactly the same way to generate simulations. The salient difference between the two methods is that GSSM uses weights extracted directly from $S(\cdot)$, which are guaranteed to be nonnegative. Like circulant embedding, simulation via GSSM is computationally appealing; however, in general it does *not* yield exact simulations. Better approximations to exact simulations can be achieved at the expensive of larger random samples and additional time to generate each simulation, which might compromise its appeal.

Turning now to case (ii), suppose that $\hat{S}(\cdot)$ is a nonparametric SDF estimate based upon a realization of X_0, X_1, \dots, X_{N-1} . If $\hat{S}(\cdot)$ is either (a) a direct spectral estimate, (b) a lag window estimate that uses a lag window such as the Parzen or Gaussian, (c) a basic or weighted multitaper estimate or

(d) a WOSA estimate, we can *always* use circulant embedding to generate simulations whose statistical properties are in accordance with $\hat{S}(\cdot)$. The weights required for circulant embedding are gotten by evaluating $\hat{S}(\cdot)$ over a grid of frequencies twice as finely spaced as that of the Fourier frequencies (an adjustment might be needed to take into account the sampling interval Δ_t ; see Section 11.4 for details). On the other hand, if $\hat{S}(\cdot)$ is a parametric AR SDF estimate and if the estimated AR coefficients used to formulate $\hat{S}(\cdot)$ correspond to those of a causal AR process (the Yule–Walker method guarantees this, as does Burg’s algorithm in practice), then we can generate simulations using the same procedure used in case (i) (we just use the estimated AR parameters as substitutes for the parameters that would be available for a fully specified AR process). This simulation procedure uses a sample of size N drawn typically from a standard Gaussian distribution. Motivated by bootstrapping, we can alternatively use samples drawn (with replacement) from normalized observed prediction errors associated with the time series behind $\hat{S}(\cdot)$ (Swanepoel and Van Wyk, 1986; see the discussion leading up to Equation (618b) for details). If the estimated AR coefficients do not correspond to those of a causal AR process, we can revert to GSSM to generate simulations that are approximately consistent with $\hat{S}(\cdot)$.

Finally, we note that the simulation techniques that address cases (i) and (ii) typically work best when our goal is to generate realizations from a Gaussian stationary process (if we deviate from this goal, in general we no longer have control over the marginal distribution of the simulations). If we want to simultaneously control the ACVS and the marginal probability density function (PDF) of the simulated series, we face challenges that can be addressed in two ways. First, under certain conditions, we can generate an independent and identically distributed (IID) noise sequence and manipulate it to produce a simulated series that both obeys a desired ARMA structure and matches at least certain aspects of a desired PDF (mean, variance, skewness and kurtosis). Second, again under certain conditions, we can filter IID Gaussian noise and then subject it to a zero-memory nonlinear (ZMNL) transform to yield a simulated series with a tractable ACVS and a tractable (and fully specified) PDF (details are in Section 11.7).

11.9 Exercises

[11.1] Consider a Gaussian AR(4) process defined by Equation (35a), namely,

$$X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9238X_{t-4} + \epsilon_t = \sum_{j=1}^4 \phi_{4,j}X_{t-j} + \epsilon_t,$$

where $\{\epsilon_t\}$ is Gaussian white noise with zero mean and variance $\sigma_4^2 = 0.002$ (hence $\sigma_4 \doteq 0.04472136$). Assuming for convenience that $N \geq 5$, an adaptation of Equation (597a) for the case $p = 4$ yields

$$\begin{aligned} X_0 &= \sigma_0 Z_0 \\ X_1 &= \phi_{1,1}X_0 + \sigma_1 Z_1 \\ X_2 &= \sum_{j=1}^2 \phi_{2,j}X_{2-j} + \sigma_2 Z_2 \\ X_3 &= \sum_{j=1}^3 \phi_{3,j}X_{3-j} + \sigma_3 Z_3 \\ X_t &= \sum_{j=1}^4 \phi_{4,j}X_{t-j} + \sigma_4 Z_t, \quad t = 4, \dots, N-1, \end{aligned}$$

which gives us a recipe for manipulating realizations of Z_0, Z_1, \dots, Z_{N-1} (a random sample of Gaussian RVs with zero means and unit variances) to produce what can be regarded as a realization of X_0, X_1, \dots, X_{N-1} . Show that, to make this scheme work for the AR(4) process of interest

here, we need to set

$$\sigma_0 \doteq 1.234\,274\,92$$

$$\sigma_1 \doteq 0.861\,043\,15, \quad \phi_{1,1} \doteq 0.716\,477\,21$$

$$\sigma_2 \doteq 0.164\,409\,13, \quad \phi_{2,1} \doteq 1.419\,772\,21, \quad \phi_{2,2} \doteq -0.981\,601\,36$$

$$\sigma_3 \doteq 0.116\,803\,96, \quad \phi_{3,1} \doteq 2.110\,574\,98, \quad \phi_{3,2} \doteq -1.980\,767\,23, \quad \phi_{3,3} \doteq 0.703\,750\,83.$$

- [11.2] (a) In the last part of C&E [2] for Section 11.1, we discussed a procedure described in McLeod and Hipel (1978) for generating realizations of a zero mean Gaussian ARMA(p, q) process $\{X_t\}$ with stationary initial conditions. Work out the details of their procedure for an ARMA(1,1) process by determining the entries of the covariance matrix Σ_{p+q} of Equation (600a). As noted in the C&E, generating the stationary initial conditions requires forming the Cholesky decomposition of Σ_{p+q} , i.e., finding a lower triangular matrix C_{p+q} such that $C_{p+q}C_{p+q}^T = \Sigma_{p+q}$. Determine C_{p+q} and describe how to use it to generate the required stationary initial conditions. Finally state how these conditions can be used in conjunction with the defining equation for an ARMA(1,1) process to generate a realization of X_0, X_1, \dots, X_{N-1} . Hint: assuming $a \neq 0$, the Cholesky decomposition of a symmetric positive semidefinite matrix of the form

$$\mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \text{ is } \mathbf{A} = \begin{bmatrix} \sqrt{a} & 0 \\ \frac{b}{\sqrt{a}} & \sqrt{c - \frac{b^2}{a}} \end{bmatrix} \begin{bmatrix} \sqrt{a} & \frac{b}{\sqrt{a}} \\ 0 & \sqrt{c - \frac{b^2}{a}} \end{bmatrix}.$$

- (b) Following Exercise [598], we discussed a different procedure – as described in Kay (1981b) – for generating realizations of an ARMA(p, q) process with stationary initial conditions. Work out the details of this procedure for the ARMA(1,1) case. Comment briefly on how it compares with the procedure formulated in part (a).
- (c) As one check to verify that the procedures of parts (a) and (b) yield equivalent simulations, show that expressions for $\text{var}\{X_0\}$ that can be deduced from the two procedures are in agreement.
- [11.3] Equation (597a) describes a procedure advocated by Kay (1981b) for simulating a portion X_0, X_1, \dots, X_{N-1} of a zero mean Gaussian AR(p) process. The procedure involves unraveling prediction errors associated with best linear predictors, with the coefficients for the predictors being provided by the Levinson–Durbin recursions described in Section 9.4. No matter what the model order p is, each simulated series requires a random sample of length N from a standard Gaussian distribution. By contrast, the procedures described in Section 11.1 for simulating a portion of length N from a zero mean MA(q) process require a random sample of length $N + q$. Focusing for simplicity on the MA(1) case, explore an approach similar to Kay’s to simulate a portion X_0, X_1, \dots, X_{N-1} of a zero mean Gaussian MA(1) process, with the goal being to use just N realizations of a standard Gaussian RV as opposed to $N + 1$ (the discussion at the beginning of Section 9.4 notes that the Levinson–Durbin recursions are not restricted to AR(p) processes, but rather can be used to form the coefficients for the best linear predictors of other stationary processes). With this same goal in mind, explore also adaptation of the approach of McLeod and Hipel (1978) that is discussed in C&E [2] for Section 11.1. Comment briefly on how the two adapted approaches compare with the procedures described in Section 11.1 for simulating an MA(1) process.
- [11.4] Consider an AR(1) process $\{X_t\}$ dictated by the equation

$$X_t = \phi_{1,1}X_{t-1} + \epsilon_t,$$

where $0 < |\phi_{1,1}| < 1$, and $\{\epsilon_t\}$ is Gaussian white noise with mean zero and variance σ_1^2 (we take the sampling interval Δ_t to be unity). As noted by Equation (507a), the ACVS for this process is given by $s_\tau = \phi_{1,1}^{|\tau|} \sigma_1^2 / (1 - \phi_{1,1}^2)$.

- (a) Given a realization of a portion Z_0, Z_1, \dots, Z_{N-1} of Gaussian white noise with zero mean and unit variance, show how to create a realization of X_0, X_1, \dots, X_{N-1} using the exact method described by Equation (597a) (i.e., develop a recipe similar to the one given in Exercise [597] for the AR(2) process of Equation (34)).

- (b) Show that the weights $\{S_k\}$ given by Equation (601a) and used by the circulant embedding method to create a realization of length N from $\{X_t\}$ take the form

$$S_k = \frac{\sigma_1^2 (1 - (-1)^k \phi_{1,1}^N)}{1 - 2\phi_{1,1} \cos(\pi k/N) + \phi_{1,1}^2} \quad k = 0, 1, \dots, 2N - 1. \quad (640a)$$

Argue that $S_k > 0$ (thus the conditions for the circulant embedding method to work hold for $\{X_t\}$).

- (c) Consider the case $\phi_{1,1} = 0.9$, $\sigma_1^2 = 1 - \phi_{1,1}^2$ and $N = 10$, for which the ACVS at the maximum lag in the circularized sequence of Equation (601b) is $s_{10} = \phi_{1,1}^{10} \doteq 0.3487$. Suppose that we replace s_{10} in the sequence with

$$\gamma \stackrel{\text{def}}{=} \frac{2\phi_{1,1}^N + \phi_{1,1} - 1}{\phi_{1,1} + 1} \doteq 0.3144 \quad (640b)$$

and then generate weights, say \tilde{S}_k , in keeping with Equation (601a); i.e.,

$$\tilde{S}_k = \sum_{\tau=0}^{N-1} s_\tau e^{-i2\pi \tilde{f}_k \tau} + \gamma e^{-i2\pi \tilde{f}_k N} + \sum_{\tau=N+1}^{2N-1} s_{2N-\tau} e^{-i2\pi \tilde{f}_k \tau}.$$

How do the \tilde{S}_k weights compare with the S_k weights? (Once this question is answered, an extension to this exercise is to explain how γ came about!)

- [11.5] Exercise [602] notes that the simulations $\{Y_{\mathcal{R},t}\}$ and $\{Y_{\mathcal{I},t}\}$ generated by the circular embedding method have ACVSs that agree with that of the desired time series $\{X_t\}$ for lags satisfying $0 \leq \tau \leq N$, but not necessarily for lags $\tau > N$. For the AR(2) process that is the subject of Figure 603, but now setting $N = 20$ rather than 1024, determine how well the ACVSs for $\{Y_{\mathcal{R},t}\}$ and $\{Y_{\mathcal{I},t}\}$ agree with that for $\{X_t\}$ at lags $\tau > 20$.
- [11.6] Consider the AR(4) process of Equation (35a) (its ACVS $\{s_\tau\}$ can be computed using the procedure described surrounding Equation (508b)). Suppose we want to generate simulations of length $N = 128$ using the circulant embedding method.
- Calculate the required $2N = 256$ weights S_k as per Equation (601a), noting that some of the weights are negative. Next calculate the $2N' = 512$ weights associated with a desired length of 256, i.e., $N' = 2N$ rather than N . Are all 512 of these weights nonnegative?
 - Reconsider the 256 weights S_k computed in part (a) for the desired length $N = 128$, and define $\tilde{S}_k = (S_k)_+$ (i.e., the maximum of S_k and 0). If we use \tilde{S}_k as a substitute for S_k in Equation (601c) and then proceed as usual with the circulant embedding method, show that the resulting simulations $Y_{\mathcal{R},0}, Y_{\mathcal{R},1}, \dots, Y_{\mathcal{R},127}$ and $Y_{\mathcal{I},0}, Y_{\mathcal{I},1}, \dots, Y_{\mathcal{I},127}$ are two independent realizations from stationary processes with a common ACVS, say $\{\tilde{s}_\tau\}$, but that this ACVS is not in perfect agreement with $\{s_\tau\}$ at lags $\tau = 0, 1, \dots, 128$. Use $\left[\frac{1}{128} \sum_{\tau=1}^{128} (\tilde{s}_\tau/\tilde{s}_0 - s_\tau/s_0)^2\right]^{1/2}$ to ascertain the degree of imperfection. Comment upon your findings.
- [11.7] Show that s'_τ as defined in Equation (606a) is real-valued.
- [11.8] Let U_t be as defined by Equation (606c), and let $U_{\mathcal{R},t}$ and $U_{\mathcal{I},t}$ be its real and imaginary components. Show that both $\{U_{\mathcal{R},t} : t \in \mathbb{Z}\}$ and $\{U_{\mathcal{I},t} : t \in \mathbb{Z}\}$ are Gaussian stationary processes with zero mean and ACVS $\{s'_\tau\}$, where s'_τ is defined by Equation (606a). Also show that each RV in $\{U_{\mathcal{R},t}\}$ is independent of each RV in $\{U_{\mathcal{I},t}\}$. Hint: study the solution to Exercise [602].
- [11.9] Verify the contents of Figure 607 and Table 607.
- [11.10] Consider again the AR(1) process $\{X_t\}$ that is the focus of Exercise [11.4].
- Assuming that $N' = N$ so that f'_k is set to $k/(2N)$, derive an expression for the weights $\{S(f'_k)\}$ needed for GSSM (these are used in Equation (606b)). Compare these weights with the weights $\{S_k\}$ of Equation (640a) that are used in circulant embedding by deriving an expression for $\max_k |S(f'_k) - S_k|$. Finally, comment upon the following question: while

circulant embedding is an exact simulation method for simulating AR(1) time series, can the same be said for GSSM?

- (b) In the paragraph containing Equation (609), we illustrated GSSM by generating simulated series of length $N = 1024$ from the AR(2) process of Equation (34). We found that, although GSSM is in theory not an exact method, when numerical accuracy comes into play, it is computationally indistinguishable from circulant embedding (an exact method). To lend credence to this finding, consider the following. As noted in the discussion surrounding Equation (177a), the dynamic range $10 \log_{10} (\max_f S(f) / \min_f S(f))$ is a crude measure of the complexity of an SDF. The AR(2) SDF is such that $10 \log_{10} (\max_f S(f)) \doteq 7.5$ dB and $10 \log_{10} (\min_f S(f)) \doteq -7.0$ dB, so its dynamic range is 14.5 dB. Determine the values of $\phi_{1,1} > 0$ and σ_1^2 that are needed so that the AR(1) SDF has the same maximum and minimum values as the AR(2) and hence the same dynamic range. For this AR(1) process and with N set to 1024, use the expression you derived in part (a) to compute $\max_k |S(f'_k) - S_k|$. Does this computation lend credence to the apparent numerical agreement between the circulant embedding weights and the GSSM weights we found in the AR(2) case?
- [11.11] At the end of Section 11.2, we illustrated the circulant embedding method by generating simulated series of length $N = 1024$ from the AR(2) process of Equation (34) – see Figure 603 and the discussion surrounding it. Replicate what was done there, but now using the AR(4) process of Equation (35a); in particular, generate the equivalent of Figure 603. We also used the AR(2) process to illustrate GSSM (see the paragraph containing Equation (609)), but did not show the equivalent of Figure 603 (such a figure was not needed because it turned out to be visually identical to Figure 603). Replicate what was done there, but now using the AR(4) process; if need be, generate the equivalent of Figure 603. Comment upon your findings.
- [11.12] For each of the four AR(2) time series of length $N = 1024$ shown in Figure 34, compute the periodogram over the grid of frequencies $\tilde{f}_k = k/(2N)$, $k = 0, 1, \dots, 2N - 1$ (the series are downloadable from the “Data” part of the website for the book). Based upon each of the four periodograms, use the circulant embedding method to generate a simulated series whose theoretical SDF is equal to the periodogram (this method gives two simulated series, but just pick one of them for subsequent use). Compute the periodograms for each of the four simulated series. Compare these periodograms to the corresponding periodograms for the four actual AR(2) time series. Repeat this exercise using the four AR(4) series in Figure 34. Comment upon your findings.
- [11.13] (a) Show that $X_{\phi,t}$ of Equation (616a) is real-valued.
 (b) Verify that $X_{\phi,t}$ of Equation (616a) can be reexpressed as in Equation (616b).
 (c) Argue that, when N is odd, $\{X_{\phi,t}\}$ is a harmonic process as defined by Equation (35c) (but formulated as per Equation (35d)), whereas, when N is even, $\{X_{\phi,t}\}$ is *not* a harmonic process in general. Hint: study Exercise [37] and its solution.
- [11.14] Consider the causal ARMA(1,1) process $X_t = \phi X_{t-1} + \epsilon_t + \vartheta \epsilon_{t-1}$, which is the focus of Exercise [632b]. Assume that $\phi + \vartheta \neq 0$, and note that causality implies $|\phi| < 1$. Starting from Equation (632a), show that $|\sqrt{\delta_1(X_t)}| < |\sqrt{\delta_1(\epsilon_t)}|$ for this process. Hint: study the proof of Equation (632e) in Exercise [632b].

