Chapter 8

Time Series

So far we have considered cases of samples where observations are independent. In many statistical applications the observations are dependent. Time series, stochastic process, Markov chain, Brownian motion, mixing, weak dependence and long-memory are just a few examples of the terminology used to describe dependent observations. Dependency may create dramatic complications in statistical analysis that should be understood and taken into account. It is worthwhile to note that, in a number of practically interesting cases, dependent observations may be considered as a modification of independent ones, and this is exactly how many classical stochastic processes are defined and/or generated. Further, the dependency will allow us to test the limits of the proposed E-estimation methodology.

Dependent observations are considered in this and next chapters. This chapter is devoted to stationary time series with the main emphasis on estimation of the spectral density and missing data, while the next chapter is primarily devoted to nonstationary dependent observations.

Sections 8.1 and 8.2 serve as the introduction to dependent observations and the spectral density, respectively. Sections 8.3 and 8.4 consider estimation of the spectral density for time series with missing observations. In particular, Section 8.4 presents a case of destructive missing. Estimation of the spectral density for censored time series is discussed in Section 8.5. Probability density estimation for dependent observations is explored in Section 8.6. Finally, Section 8.7 is devoted to the problem of nonparametric autoregression.

8.1 Discrete-Time Series

A discrete-time series (which also may be referred to as a discrete-time process, or a process, or a time series) is a set of pairs of observations (t_1, X_1), (t_2, X_2), $\dots$, (t_n, X_n) where each response X_s has been recorded at a specific time t_s, and traditionally $t_1 < t_2 < \cdots < t_n$. The setting resembles regression, only now the main issue is that responses X_s, $s = 1, 2, \dots$ may be dependent and the main task is to explore the relationship between the responses and understand how to solve practical problems that involve dependent observations. (In general t_i may not necessarily be a time but a location like in spatial geostatistical data or the order of insurance claim; while there are some specifics in those cases, the general methodology remains the same.) A typical feature of a time series is that predictors t_l are equidistant integers. Then, without loss of generality, we may set $t_l = l$, and the corresponding time series is completely described by the responses $\{X_l,\, l = 1, 2, \dots\}$, which may be treated as a sequence of regular observations in time, and this explains why such a sequence is called a *time series*. Many practical examples are indeed sequences in time, but there are plenty of other examples; for instance, data may be collected in space. In the latter case the data are often referred to as *spatial data*, and there is even a special branch in statistics, known as geostatistics, that is primarily concerned with the analysis of such data. In this chapter, for the sake of clarity, we shall use only time series terminology and assume that data are collected sequentially in time.

A time series is called *strictly stationary* (or *stationary*) if the joint distribution of $(X_{s_1}, X_{s_2}, \ldots, X_{s_m})$ and $(X_{s_1+k}, X_{s_2+k}, \ldots, X_{s_m+k})$ are the same for all sets $(s_1, \ldots, s_m)$ and all integers k, m. In other words, a shift in time does not change the joint distribution and thus the time series is stationary in time. Note that no assumption about moments is made, and for instance a time series of independent realizations of a Cauchy random variable is a strictly stationary time series.

A time series $\{X_t\}$ is called *zero-mean* if $\mathbb{E}\{X_t\} = 0$ for all t. Note that a zero-mean time series assumes existence of the first moment, but no other assumptions about moments or the distribution is made.

A time series $\{X_t\} := \{\ldots, X_{-1}, X_0, X_1, \ldots\}$ is called *second-order stationary* time series if: (i) $E\{X_t^2\} < \infty$ for all t, that is, the second moment is finite; (ii) $E\{X_t\} = \mu$ for all t, that is, the expectation is constant; (iii) the *autocovariance function* $\gamma^X(l, s) := E\{(X_l - \mu)(X_s - \mu)\}$ satisfies the relation $\gamma^X(l, s) = \gamma^X(l + k, s + k)$ for all integers l, s, and k, that is, a translation in time does not affect the autocovariance function. The property (iii) implies that $\gamma^X(l, s) =: \gamma^X(l - s) = \gamma^X(s - l)$. To see this just set $k = -s$ and $k = -l$. Thus a zero-mean and second-order stationary time series is characterized by its autocovariance function $\gamma^X(k)$ at the *lag* k, and further there is a nice relation $\gamma^X(0) = \mathbb{E}\{X_t^2\} = \mathbb{V}(X_t)$ which holds for all t. Also note that no assumptions about higher moments is made for a second-order stationary time series.

Now let us comment about estimation of the mean of a second-order stationary time series $\{Y_t\} := \{\mu + X_t\}$ where $\{X_t\}$ is a zero-mean and second-order stationary time series with the autocovariance function $\gamma^X(t) := \mathbb{E}\{(X_t - \mathbb{E}\{X_t\})(X_0 - \mathbb{E}\{X_0\})\} = \mathbb{E}\{X_t X_0\}$. Suppose that we observe a realization $Y_1, \ldots, Y_n$ of $\{Y_t\}$. First of all we note that

$$\mathbb{E}\{Y_t\} = \mathbb{E}\{\mu + X_t\} = \mu + \mathbb{E}\{X_t\} = \mu. \tag{8.1.1}$$

This formula immediately yields the following sample mean estimator of μ,

$$\bar{\mu} := n^{-1} \sum_{l=1}^{n} Y_l. \tag{8.1.2}$$

Expectation of the sum of random variables is always (regardless of dependence between that variables) the sum of expectations of the random variables, and hence the sample mean estimator (8.1.2) is unbiased. Let us also note that $\gamma^X(j) = \gamma^Y(j)$ for all j.

What can be said about variance of the sample mean estimator (8.1.2)? So far we have been dealing only with independent observations and it is of interest to understand how to deal with dependent variables. There is a special technique for dealing with sums of dependent variables and we should learn it. Write,

$$\mathbb{V}(\bar{\mu}) = \mathbb{E}\{[n^{-1} \sum_{l=1}^{n} Y_j - \mu]^2\}$$

$$= \mathbb{E}\{[n^{-1} \sum_{l=1}^{n} (Y_l - \mu)]^2\} = \mathbb{E}\{[n^{-1} \sum_{l=1}^{n} X_l]^2\}. \tag{8.1.3}$$

The squared sum in the right side of (8.1.3) may be written as a double sum, and we continue (8.1.3),

$$\mathbb{V}(\bar{\mu}) = \mathbb{E}\{[n^{-1} \sum_{l=1}^{n} X_l]^2\} = \mathbb{E}\{n^{-2} \sum_{l,t=1}^{n} X_l X_t\}$$

$$= n^{-2} \sum_{l,t=1}^{n} \gamma^X(l, t) = n^{-1}[n^{-1} \sum_{l=1}^{n} \{\sum_{t=1}^{n} \gamma^X(l - t)\}]. \tag{8.1.4}$$

Note that only in the last equality we used the second-order stationarity of $\{X_t\}$.

Relation (8.1.4) is the result that we need. Let us consider several possible scenarios. If the observations are independent, then $\gamma^X(l) = 0$ for any $l \neq 0$, and (8.1.4) together with $\gamma^X(0) = \mathbb{V}(X_1)$ imply the familiar formula $\mathbb{V}(\bar{\mu}) = n^{-1}\mathbb{V}(X_1)$. Similarly, if $\sum_{l=0}^{\infty} |\gamma^X(l)| < c < \infty$ then $\mathbb{V}(\bar{\mu}) < cn^{-1}$ and we again have the parametric rate of the variance convergence. In the latter case the stochastic process may be referred to as a *short-memory* time series. However, if the sum $\sum_{l=0}^{n} \gamma^X(l)$ diverges, then we lose the rate n^{-1}. For instance, consider the case of a *long-memory* time series when $\gamma^X(t)$ is proportional to $|t|^{-\alpha}$, $0 < \alpha < 1$. Then the sum in the curly brackets on the right side of (8.1.4) is proportional to $n^{1-\alpha}$ and the variance is proportional to $n^{-\alpha}$. This is a dramatic slowing down of the rate of convergence caused by dependence between observations. The above-explained phenomenon sheds light on complexity of dealing with stochastic processes.

The simplest zero-mean and second-order stationary time series is a process in which the random variables $\{X_t\}$ are uncorrelated (that is, $\gamma^X(t) = 0$ for $t \neq 0$) and have zero mean and unit variance. Let us denote this time series as $\{W_t\}$ and call it a *standard (discrete time) white noise*. A classical example is a time series of independent standard Gaussian random variables, which is the white noise that will be used in the following simulations, and we call it a *standard Gaussian white noise*.

In its turn, a white noise allows us to define a wide variety of dependent second-order stationary and zero-mean processes via a set of linear difference equations. This leads us to the notion of an *autoregressive moving average process of orders p and q*, an ARMA(p, q) process for short. By definition, the process $\{X_t, t = \ldots, -1, 0, 1, \ldots\}$ is said to be an ARMA(p, q) process if $\{X_t\}$ is zero-mean and second-order stationary, and for every t

$$X_t - a_1 X_{t-1} - \cdots - a_p X_{t-p} = \sigma(W_t + b_1 W_{t-1} + \cdots + b_q W_{t-q}), \tag{8.1.5}$$

where $\{W_t\}$ is a standard white noise, $\sigma > 0$, the orders p and q are nonnegative integers, and $a_1, \ldots, a_p, b_1, \ldots, b_p$ are real numbers. For the case of a Gaussian white noise we shall refer to the corresponding ARMA process as a *Gaussian ARMA process*.

Two particular classical examples of an ARMA process are a *moving average* MA(q) process, which is a moving average of $q + 1$ consecutive realizations of a white noise,

$$X_t = \sigma(W_t + b_1 W_{t-1} + \cdots + b_q W_{t-q}), \tag{8.1.6}$$

and an *autoregressive* AR(p) process satisfying the difference equation

$$X_t - a_1 X_{t-1} - \cdots - a_p X_{t-p} = \sigma W_t. \tag{8.1.7}$$

The MA and AR processes play an important role in the analysis of time series. For instance, prediction of values $\{X_t,\ t \geq n + 1\}$ in terms of $\{X_1, \ldots, X_n\}$ is relatively simple and well understood for an autoregressive process because $\mathbb{E}\{X_t | X_{t-1}, X_{t-2}, \ldots\} = a_1 X_{t-1} + \ldots + a_p X_{t-p}$. Also, for a given autocovariance function it is simpler to find an AR process with a similar autocovariance function. More precisely, if an autocovariance function $\gamma^X(j)$ vanishes as $j \to \infty$, then for any integer k one can find an AR(k) process with the autocovariance function equal to $\gamma^X(j)$ for $|j| \leq k$. The "negative" side of an AR process is that it is not a simple issue to find a stationary solution for (8.1.7), and moreover, it may not exist. For instance, the difference equation $X_t - X_{t-1} = \sigma W_t$ has no stationary solution, and consequently there is no AR(1) process with $a_1 = 1$. A thorough discussion of this issue is beyond this short introduction, and in what follows a range for the coefficients that "keeps us out of trouble" will be always specified.

The advantages of a moving average process are its simple simulation, the given expression for a second-order stationary solution, and that it is very close by its nature to white noise, namely, while realizations of a white noise are uncorrelated, realizations of an MA(q)

process are also uncorrelated whenever the lag is larger than q. The disadvantages, with respect to AR processes, are more complicated procedures for prediction and estimation of parameters. Thus, among the two, typically AR processes are used for modeling and prediction. Also, AR processes are often used to approximate an ARMA process.

For a time series, and specifically ARMA processes, the notion of causality (future independence) plays an important role. The idea is that for a causal ARMA process $\{X_t\}$ (or more specifically, a causal process with respect to an underlying white noise $\{W_t\}$) it is quite natural to expect that an ARMA time series $\{X_t\}$ depends only on current and previous (but not future!) realizations of the white noise. Thus, we say that an ARMA process $\{X_t\}$ generated by a white noise $\{W_t\}$ is *causal* if $X_t = \sum_{j=0}^{\infty} c_j W_{t-j}$, where the coefficients c_j are absolutely summable. Clearly, MA(q) processes are causal, but not all AR(p) processes are; for instance, a stationary process corresponding to the difference equation $X_t - 2X_{t-1} = W_t$ is not causal. We shall not elaborate more on this issue and only note that in what follows we are considering simulations of Gaussian ARMA(1,1) processes corresponding to the difference equation $X_t - aX_{t-1} = \sigma(W_t + bW_{t-1})$ with $|a| < 1$ and $-a \neq b$. It may be directly verified that for such a this equation has a stationary and causal solution $X_t = \sigma W_t + \sigma(a+b)\sum_{j=1}^{\infty} a^{j-1} W_{t-j}$.

As we know from our discussion of (8.1.4), for a statistical inference it is important to know how fast autocovariance function $\gamma^X(k)$ decreases in k. Introduce a class of autocovariance functions

$$\mathcal{A}(Q, q, \beta, r) = \{\gamma^X : \ 0 < q < \gamma^X(0) \le Q < \infty,$$

$$|\gamma^X(k)| \le Q(k+1)^{\beta} e^{-rk}(1 + o_k(1)), \ k = 1, 2, \ldots\}. \tag{8.1.8}$$

Recall that $o_k(1)$ denote generic sequences such that $o_k(1) \to 0$ as $k \to \infty$. The importance of class (8.1.8) is explained by the fact that the autocovariance function of a causal ARMA(p, q) time series decreases exponentially (compare with the analytic class (2.1.2)). Furthermore, the parameter r in (8.1.8) is the logarithm of the minimal modulus of zeroes of the autoregressive polynomial, and an ARMA process is causal if and only if r is positive. Furthermore, for any quartet $\{Q, q, \beta, r\}$ there exists a causal ARMA(p, q) process such that in (8.1.8) we get equalities. Of course, class (8.1.8) includes not only spectral densities of ARMA time series. For instance, note that covariance of the product of two independent zero-mean time series is the product of covariances of these series. As a consequence if a zero-mean and second order stationary time series is multiplied by a time series from class (8.1.8) (in particular by an ARMA time series), then the product also belongs to class (8.1.8).

Apart of linear (with respect to an underlying white noise) processes, in many applications an underlying time series may not be linear. For instance, many interesting processes may be modeled by a nonparametric autoregressive process $X_t = q(X_{t-1}, \ldots, X_{t-p}) + \sigma(X_{t-1}, \ldots, X_{t-p})W_t$ where $q(\cdot)$ and $\sigma(\cdot)$ are p-variate functions. We will continue discussion of this process in Section 8.7.

So far we have been using a time domain (dynamic) approach to describe a process via evolution of a stochastic process in time. Another approach, considered in the next section, is to look at a second-order stationary process in the spectral (frequency) domain. And yet another approach is to explore the dependence between events as the time between them increases. The theory that uses this approach is called the *mixing* theory. Let us consider a stationary time series $\{X_t\}$ and introduce a class $\mathcal{M}_{-\infty}^{k}$ of all possible events of interest generated by variables $\{\ldots, X_{k-1}, X_k\}$ and a class $\mathcal{M}_{k}^{\infty}$ of all possible events of interest generated by variables $\{X_k, X_{k+1}, \ldots\}$ (a rigorous definition of these classes can be found in references mentioned in the Notes where the class is called a sigma-algebra). Consider a particular t and two events $G_t \in \mathcal{M}_{-\infty}^{t}$ and $G^{t+s} \in \mathcal{M}_{t+s}^{\infty}$. If the events are independent,

then we know that

$$\mathbb{P}(G_t \cap G^{t+s}) - \mathbb{P}(G_t)\mathbb{P}(G^{t+s}) = 0. \tag{8.1.9}$$

Note that due to stationarity we can set $t = 0$, in other words if (8.1.9) holds for a particular $t = t_0$, then it holds for all t. Then the underlying idea of the mixing theory is to describe the dependence between separated in time events via deviation of the left side of (8.1.9) from zero. For instance, for a stationary time series $\{X_t\}$ we can introduce a *mixing coefficient*

$$\alpha^X(s) := \sup_{G_0 \in \mathcal{M}^0_{-\infty}, G^s \in \mathcal{M}^\infty_s} \{|\mathbb{P}(G_0 \cap G^s) - \mathbb{P}(G_0)\mathbb{P}(G^s)|\}, \;\; s > 0. \tag{8.1.10}$$

Let us present several general properties and examples of mixing coefficients. Mixing coefficient $\alpha^X(s)$ is either positive or equal to zero, and not increasing in s. For the case of a stationary time series $\{X_t\}$ of independent variables we have $\alpha^X(s) = 0$. Another important example is an *m-dependent* series satisfying $\alpha^X(s) = 0$ for $s > m$. The meaning of m-dependence is that variables, separated in time for more than m time-units, are independent. Another class of time series, often considered in the mixing theory, is when for $\tau > 8$

$$\sum_{s=0}^{\infty} (s+1)^{\tau/2-1} \alpha^X(s) \le Q < \infty. \tag{8.1.11}$$

Further, an important Kolmogorov-Rosanov result for a Gaussian stationary time series $\{X_t\}$ is that the covariance $\gamma^X(s)$ is of the same order as the mixing coefficient $\alpha^X(s)$, and this result bridges the classical second-order stationary time series theory with the mixing theory.

Let us also mention several classical results of the mixing theory that allow us to analyze expectations. Consider a stationary time series $\{X_t\}$, a random variable Y_0 which is a function of $\{\ldots, X_{-1}, X_0\}$, and a random variable Z_s which is a function of $\{X_s, X_{s+1}, \ldots\}$, $s > 0$. For a time series of independent variables we have $\mathbb{E}\{Y_0 Z_s\} - E\{Y_0\}E\{Z_s\} = 0$, and for an m-dependent series we have $\mathbb{E}\{Y_0 Z_s\} - E\{Y_0\}E\{Z_s\} = 0$ for $s > m$. Further, suppose that $\mathbb{E}\{|Y_0|^{2+\delta}\} < \infty$ and $\mathbb{E}\{|Z_s|^{2+\delta}\} < \infty$ for some $\delta > 0$, then there exists a finite constant c such that

$$|\mathbb{E}\{Y_0 Z_s\} - E\{Y_0\}E\{Z_s\}| \le c[\alpha^X(s)]^{\delta/(2+\delta)}. \tag{8.1.12}$$

Further, if $\mathbb{P}(|Y_0| > c_1) = 0$ and $\mathbb{P}(|Z_s| > c_2) = 0$ then

$$|\mathbb{E}\{Y_0 Z_s\} - E\{Y_0\}E\{Z_s\}| \le 4c_1 c_2 \alpha^X(s). \tag{8.1.13}$$

These are types of results that allow us to develop statistical inference in problems with stationary dependent variables, more results (including rigorous measure-theoretical formulations) may be found in the references mentioned in the Notes.

The reader may also recall that a discrete time Markov chain is another classical approach for modeling and analysis of dependent observations. It will be briefly discussed in Section 8.3.

8.2 Spectral Density and Its Estimation

Let $\{X_t\}$ be a zero-mean and second-order stationary time series with the autocovariance function $\gamma^X(j) := E\{X_t X_{t+j}\}$. The second-order properties of this time series are completely described by its autocovariance function, or, equivalently, under mild conditions (for instance, a sufficient condition is $\sum_{j=-\infty}^{\infty} |\gamma^X(j)| < \infty$), by its Fourier transform, which is

called the *spectral density* function,

$$g^X(\lambda) := (2\pi)^{-1}\sum_{j=-\infty}^{\infty}\gamma^X(j)\cos(j\lambda) \tag{8.2.1}$$

$$= (2\pi)^{-1}\gamma^X(0)+\pi^{-1}\sum_{j=1}^{\infty}\gamma^X(j)\cos(j\lambda),\quad -\pi<\lambda\le\pi. \tag{8.2.2}$$

Here the frequency λ is in units radians/time, and to establish the equality (8.2.2) we used the relation $\gamma^X(-j)=\gamma^X(j)$. The spectral density is symmetric in λ about 0, i.e., the spectral density is an even function. Thus, it is customary to consider a spectral density on the interval $[0,\pi]$. The spectral density is also a nonnegative function (like the probability density), and this explains why it is called a density.

One of the important applications of the spectral density is searching for a deterministic periodic component (often referred to as a seasonal component) in nonstationary time series. Namely, a peak in $g^X(\lambda)$ at frequency λ^* indicates a possible periodic phenomenon with period

$$T^*=\frac{2\pi}{\lambda^*}. \tag{8.2.3}$$

This formula explains why spectral domain analysis is the main tool in searching after the period of a seasonal component, and the estimator will be discussed in Section 9.3.

Let us explain how the spectral density may be estimated using our E-estimation methodology. But first let us pause for a moment and stress the following important remark. By its definition, spectral density is a *cosine series*, and this is an example where the basis is chosen not due to its convenience, as we did in the cases of E-estimation of the probability density and the regression, but due to definition of the estimand. In other words, spectral density estimation is the most appealing example of using a series approach and the cosine basis. Now let us return to estimation of the spectral density.

Denote by $X_1,\dots,X_n$ the realization of a second-order stationary and zero-mean time series. The classical *empirical autocovariance estimator* is defined as

$$\tilde\gamma^X(j):=n^{-1}\sum_{l=1}^{n-j}X_lX_{l+j},\quad j=0,1,\dots,n-1, \tag{8.2.4}$$

while the sample mean estimator is

$$\hat\gamma^X(j):=(n-j)^{-1}\sum_{l=1}^{n-j}X_lX_{l+j},\quad j=0,1,\dots,n-1. \tag{8.2.5}$$

Note that in the empirical autocovariance the divisor n is not equal to the number $n-j$ of terms in the sum, and hence it is a biased estimator. On the other hand, this divisor ensures that an estimate corresponds to some second-order stationary series. For all our purposes there is no difference between using the two estimators, but it is always a good idea to check which one is used by a statistical software. In what follows, proposed E-estimators will be based on the sample mean autocovariance estimator (8.2.5), and the reason is to follow our methodology of sample mean estimation. On the other hand, many classical spectral estimators, like the periodogram discussed below, use the estimator (8.2.4).

Based on (8.2.2), if one wants to estimate a spectral density and is not familiar with basics of nonparametric estimation discussed in Chapter 2, it is natural to plug the empirical autocovariance (8.2.4) in place of unknown autocovariance. And sure enough, this approach is well known and the resulting estimator (up to the factor $1/2\pi$) is called a *periodogram*,

$$\mathcal{I}^X(\lambda):=\tilde\gamma^X(0)+2\sum_{j=1}^{n-1}\tilde\gamma^X(j)\cos(j\lambda)=n^{-1}\Big|\sum_{l=1}^{n}X_le^{-il\lambda}\Big|^2. \tag{8.2.6}$$

Here i is the imaginary unit, i.e., $i^2 := -1$, $e^{ix} = \cos(x) + i\sin(x)$, and the periodogram is defined at the so-called *Fourier frequencies* $\lambda_k := 2\pi k/n$, where k are integers satisfying $-\pi < \lambda_k \le \pi$.

Periodogram, as a tool for spectral-domain analysis, was proposed in the late nineteenth century. It has been both the glory and the curse of the spectral analysis. The glory, because many interesting practical problems were solved at a time when no computers were available. The curse, because the periodogram, which had demonstrated its value for locating periodicities (recall (8.2.3)), proved to be an erratic and inconsistent estimator. The reason for the failure of the periodogram is clear from the point of view of nonparametric curve estimation theory discussed in Chapter 2. Indeed, based on n observations, the periodogram estimates n Fourier coefficients, and this explains the erratic performance and inconsistency. Nonetheless, it is still a popular estimator.

Using the sample mean estimator (8.2.5), we may use the E-estimator of Section 2.2 for estimation of the spectral density. Of course, the theory of E-estimation was explained for the case of independent observations, but as we will see shortly, it can be extended to dependent observations. We are beginning with a simulated example which sheds light on the problem, periodogram and E-estimator, and then explore the theory.

Figure 8.1 allows us to visualize an ARMA process, its spectral density and the two above-defined estimates of the spectral density. A particular realization of the Gaussian ARMA(1, 1) time series $X_t + 0.3X_{t-1} = 0.5(W_t - 0.6W_{t-1})$ is shown in the top diagram. Note how fast observations oscillate over time. This is because here the covariance between X_t and X_{t-1} is negative. This follows from the following formula for calculating the autocovariance function of the causal ARMA(1, 1) process $X_t - aX_{t-1} = \sigma(W_t + bW_{t-1})$ with $|a| < 1$,

$$\begin{aligned} \gamma^X(0) &= \frac{\sigma^2[(a+b)^2 + 1 - a^2]}{(1-a^2)}, \quad \gamma^X(1) = \frac{\sigma^2(a+b)(1+ab)}{(1-a^2)}, \\ \gamma^X(j) &= a^{j-1}\gamma(1), \quad j \ge 2. \end{aligned} \tag{8.2.7}$$

Note that if $a > 0$ and $b > 0$, then $\gamma(1) > 0$, and a realization of the time series will "slowly" change over time. On the other hand, if $a + b < 0$ and $1 + ab > 0$ then a realization of the time series may change its sign almost every time. Thus, depending on a and b, we may see either slow or fast oscillations in a realization of an ARMA(1, 1) process. Figure 8.1 allows us to change parameters of the ARMA process and observe different interesting patterns in this pure stochastic process. In particular, to make the process slowly changing and even see interesting repeated patterns in a time series, choose positive parameters a and b.

The solid line in the bottom diagram of Figure 8.1 shows us the underlying theoretical spectral density of the ARMA(1, 1) process. As we see, because here both a and b are negative, in the spectral domain high frequencies dominate low frequencies. The formula for calculating the spectral density is $g^X(\lambda) = \sigma^2|1 + be^{i\lambda}|^2/[2\pi|1 - ae^{i\lambda}|^2]$, and it is a particular case of the following formula for a causal ARMA(p, q) process (8.1.5),

$$g^X(\lambda) = \frac{\sigma^2\Big|1 + \sum_{j=1}^q b_j e^{-ij\lambda}\Big|^2}{2\pi\Big|1 - \sum_{j=1}^p a_j e^{-ij\lambda}\Big|^2}. \tag{8.2.8}$$

The middle diagram shows us that the periodogram has a pronounced mode at frequency $\lambda^* \approx 2.6$ which, according to (8.2.3), indicates a possibility of a deterministic periodic (seasonal) component with the period which is either 2 or 3. One may see or not see such a component in the data, but thanks to the simulation we do know that there is no periodic component in the data. The reader is advised to repeat this figure and get used to reading a periodogram because it is commonly used by statistical softwares. The bottom diagram exhibits the spectral density E-estimate which correctly shows the absence of any periodic

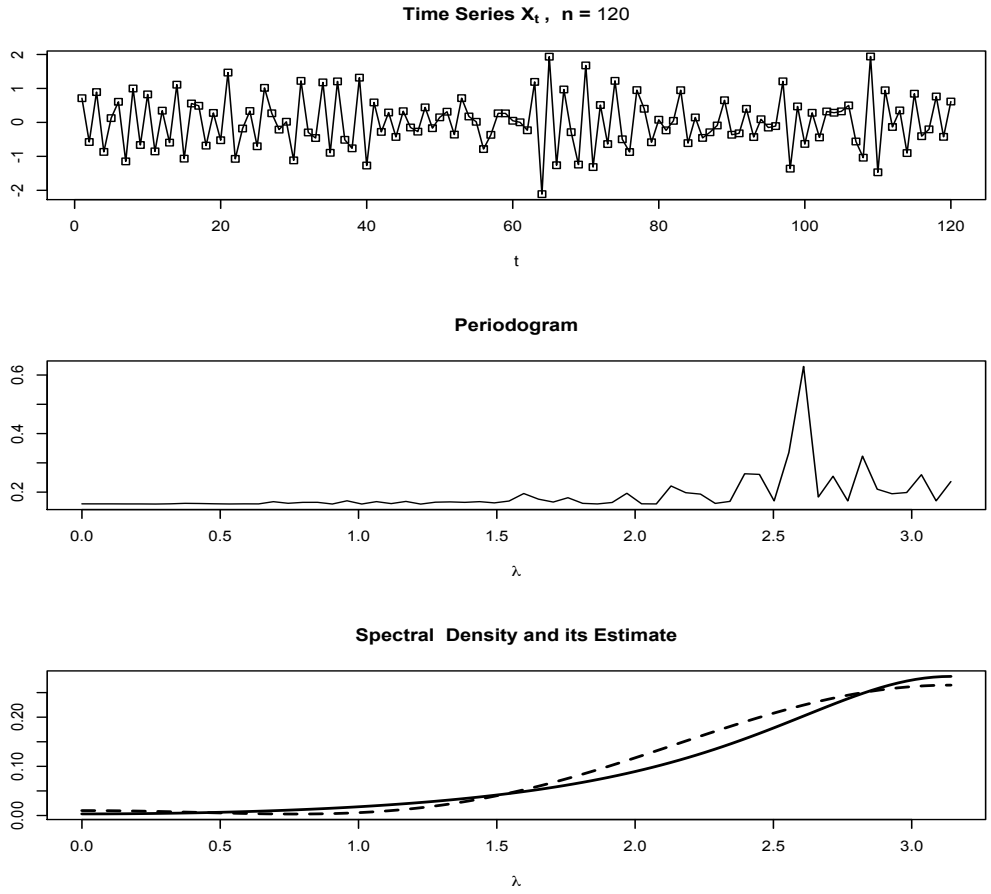

Figure 8.1 *ARMA(1,1) time series and two estimates of the spectral density. The top diagram shows a particular realization of a Gaussian ARMA(1,1) time series* $Y_t - aY_{t-1} = \sigma(W_t + bW_{t-1})$, $t = 1, 2, \ldots, n$, *where* $a = -0.3$, $b = -0.6$, $\sigma = 0.5$, *and* $n = 120$. *The middle diagram shows the periodogram. The spectral density E-estimate (the dashed line) and the underlying spectral density (the solid line) are exhibited in the bottom diagram.* {*The length n of a realization is controlled by the argument n. Parameters of simulated ARMA(1,1) process are controlled by the arguments sigma, a, and b. Use* $|a| < 1$. *All the other arguments control parameters of the E-estimator. Note that the string sp is added to these arguments to indicate that they control coefficients of the spectral density E-estimator.*} *[n = 120, sigma = 0.5, a = -0.3, b = -0.6, cJ0sp = 2, cJ1sp = 0.5, cTHsp = 4]*

(seasonal) component, and the E-estimate nicely resembles the underlying spectral density. Please pay attention to the relatively small sample size $n = 120$. Again, it is important to repeat this figure, with different sample sizes and for different ARMA processes, to get first-hand experience in spectral analysis of stationary time series.

We are finishing this section with theoretical analysis of the MISE of a series estimator. This is an interesting and technically challenging problem because a relatively simple technique of inference for a sum of independent variables no longer is applicable.

First, let us begin with an example of using the Parseval identity. We can write that

$$\int_{-\pi}^{\pi} [g^X(\lambda)]^2 d\lambda = (2\pi)^{-1}[\gamma^X(0)]^2 + \pi^{-1}\sum_{j=1}^{\infty}[\gamma^X(j)]^2. \tag{8.2.9}$$

Similarly, for a series estimator

$$\bar{g}^X(\lambda, J) := (2\pi)^{-1}\hat{\gamma}^X(0) + \pi^{-1}\sum_{j=1}^{J}\hat{\gamma}^X(j)\cos(j\lambda), \tag{8.2.10}$$

we get the following expression for its MISE,

$$\text{MISE}(\bar{g}^X(\lambda, J), g^X(\lambda)) := \mathbb{E}\{\int_{-\pi}^{\pi}[\bar{g}^X(\lambda, J) - g^X(\lambda)]^2 d\lambda\}$$

$$= \Big[(2\pi)^{-1}\mathbb{E}\{[\hat{\gamma}^X(0) - \gamma^X(0)]^2\} + \pi^{-1}\sum_{j=1}^{J}\mathbb{E}\{[\hat{\gamma}^X(j) - \gamma^X(j)]^2\}\Big] + \pi^{-1}\sum_{j>J}[\gamma^X(j)]^2. \tag{8.2.11}$$

In (8.2.11) the term in the large square brackets is the integrated variance (or simply variance) of $\bar{g}^X(\lambda, J)$, and the last term is the integrated squared bias of $\bar{g}^X(\lambda, J)$. This is a classical decomposition of the MISE (recall our discussion in Chapter 2).

Now we need to learn the technique of inference for a sum of dependent variables. In what follows we are considering $j < n$ and assume that $\{X_t\}$ is a Gaussian zero-mean and second-order stationary time series. Recall that the expectation of a sum is always the sum of expectations, and hence the mean of the sample mean autocovariance (8.2.5) is

$$\mathbb{E}\{\hat{\gamma}^X(j)\} = (n-j)^{-1}\sum_{l=1}^{n-j}\mathbb{E}\{X_l X_{l+j}\} = \gamma^X(j). \tag{8.2.12}$$

We conclude that the sample mean estimator (8.3.5) is unbiased and that the dependence does not change this nice property of a sample mean estimator. Now we are considering the variance of the sample mean estimator, and this problem is more involved because we need to learn several technical steps. Write,

$$\mathbb{V}(\hat{\gamma}^X(j)) = (n-j)^{-2}\mathbb{E}\{[\sum_{l=1}^{n-j}(X_l X_{l+j} - \gamma^X(j)]^2\}. \tag{8.2.13}$$

This is the place where we may either continue by going from the squared sum to a double sum and then do a number of calculations, or use the following nice formula. Consider a zero-mean and second-order stationary time series $\{Z_t\}$. Then, using the technique of (8.1.3)-(8.1.4) we get a formula

$$\mathbb{E}\{[\sum_{l=1}^{k} Z_l]^2\} = \sum_{l=-k}^{k}(k - |l|)\mathbb{E}\{Z_0 Z_l\}. \tag{8.2.14}$$

Note how simple, nice and symmetric the formula is, and it allows us to write down the variance of a sum of dependent variables as a sum of expectations. Using (8.2.14) in (8.2.13) we get for $j < n$,

$$\mathbb{V}(\hat{\gamma}^X(j)) = (n-j)^{-2}\sum_{l=-n+j}^{n-j}(n - j - |l|)[\mathbb{E}\{X_0 X_j X_l X_{l+j}\} - (\gamma^X(j))^2]. \tag{8.2.15}$$

Our next step is to understand how to evaluate terms $\mathbb{E}\{X_0 X_j X_l X_{l+j}\}$. In Section 8.1 several possible paths, depending on the assumption about dependency, were discussed. Here it is assumed that the time series is Gaussian, and this implies the following nice formula,

$$\mathbb{E}\{X_0 X_j X_l X_{l+j}\} = (\gamma^X(j))^2 + (\gamma^X(l))^2 + \gamma^X(l+j)\gamma^X(l-j). \tag{8.2.16}$$

Using (8.2.16) in the right side of (8.2.15) yields for $j < n$,

$$\mathbb{V}(\hat{\gamma}^X(j)) = (n-j)^2 \sum_{l=-n+j}^{n-j} (n-j-|l|)[(\gamma^X(l))^2 + \gamma^X(l+j)\gamma^X(l-j)]. \tag{8.2.17}$$

We need to add one more assumption that holds for all short-memory time series (including ARMA)

$$\sum_{l=0}^{\infty} |\gamma^X(l)| < \infty. \tag{8.2.18}$$

Note that the Cauchy-Schwarz inequality implies that $|\gamma^X(j)| \le \gamma^X(0)$, and this together with (8.2.18) yields another useful inequality

$$\sum_{l=0}^{\infty} [\gamma^X(l)]^2 \le \gamma^X(0) \sum_{l=0}^{\infty} |\gamma^X(l)| < \infty. \tag{8.2.19}$$

Using these two inequalities in (8.2.17), together with the Cauchy inequality

$$2|\gamma^X(l+j)\gamma^X(l-j)| \le \rho[\gamma^X(l+j)]^2 + \rho^{-1}[\gamma^X(l-j)]^2, \quad \rho > 0, \tag{8.2.20}$$

which is valid for any positive number ρ, we conclude that

$$\mathbb{V}(\hat{\gamma}^X(j)) = n^{-1} d(g^X)(1 + c_{n,j}), \tag{8.2.21}$$

where $c_{n,j} \to 0$ as both n and j increase in such a away that $j < J_n = o_n(1)n$, and

$$d(g^X) := [\gamma^X(0)]^2 + 2\sum_{l=1}^{\infty} [\gamma^X(l)]^2 = 2\pi \int_{-\pi}^{\pi} [g^X(\lambda)]^2 d\lambda. \tag{8.2.22}$$

In (8.2.22) the equality is valid due to (8.2.9).

Relation (8.2.21) is important because it shows us that asymptotically, given the weak dependence property (8.2.18), the variance (and the mean squared error) of the sample mean autocovariance decreases with the same classical parametric rate n^{-1} as the variance of the sample mean of independent observations. Furthermore, there is a simple expression for the coefficient of difficulty $d(g^X)$ which is proportional to the integrated squared spectral density.

Using (8.2.21) in (8.2.11) for $J = J_n$ satisfying $J_n = o_n(1)n$ and $J_n \to \infty$ as $n \to \infty$, we conclude that

$$\mathrm{MISE}(\bar{g}^X(\lambda, J_n), g^X(\lambda)) = J_n n^{-1} 2 \int_{-\pi}^{\pi} [g^X(\lambda)]^2 d\lambda (1 + o_n(1)) + \pi^{-1} \sum_{j > J_n} [\gamma^X(j)]^2. \tag{8.2.23}$$

This is a general expression for the MISE of a series estimator $\bar{g}^X(\lambda, J_n)$ defined in (8.2.10). To simplify it further, we need to add an assumption which will allow us to evaluate the second term (the ISB) on the right side of (8.2.23). As an example, let us additionally assume that the considered Gaussian time series is ARMA. Then the autocovariance function belongs to a class $\mathcal{A}(Q, q, \beta, r)$ defined in (8.1.8). For this class the assumption (8.2.18) holds, we can bound from above the sum in (8.2.23) and get the upper bound for the MISE,

$$\mathrm{MISE}(\bar{g}^X(\lambda, J_n), g^X(\lambda))$$

$$\le [2 \int_{-\pi}^{\pi} [g^X(\lambda)]^2 d\lambda J_n n^{-1} + \pi^{-1} Q^2 \sum_{j > J_n} (j+1)^{2\beta} e^{-2rj}](1 + o_n(1)). \tag{8.2.24}$$

Now we can minimize the right side of (8.2.24) with respect to cutoff J_n. We have a large choice of sequences J_n that may do this. For instance, we can choose $J_n = J_n'$ where J_n' is the largest integer smaller than $(2r)^{-1}\ln(n)(1+1/\sqrt{\ln(n)})$. This yields

$$\frac{\mathrm{MISE}(\bar g^X(\lambda, J_n'), g^X(\lambda))}{\int_{-\pi}^{\pi}[g^X(\lambda)]^2 d\lambda} \le r^{-1}\ln(n)n^{-1}(1+o_n(1)). \tag{8.2.25}$$

Note that this choice of the cutoff makes the integrated squared bias asymptotically smaller in order than the variance. This is an important property which is typical for ARMA processes. Further, this also means that in (8.2.25) we can replace the inequality on equality and get

$$\frac{\mathrm{MISE}(\bar g^X(\lambda, J_n'), g^X(\lambda))}{\int_{-\pi}^{\pi}[g^X(\lambda)]^2 d\lambda} = r^{-1}\ln(n)n^{-1}(1+o_n(1)). \tag{8.2.26}$$

Further, note that J_n' uses only parameter r of the class (8.1.8).

Asymptotic theory shows that no other estimator can improve the right side of (8.2.26) uniformly over the class (8.1.8), namely for any (not necessarily series) estimator $\check g^X(\lambda)$ the following lower bound holds,

$$\sup_{g^X\in\mathcal{A}(Q,q,\beta,r)}\Big\{\frac{\mathrm{MISE}(\check g^X(\lambda), g^X(\lambda))}{\int_{-\pi}^{\pi}[g^X(\lambda)]^2 d\lambda}\Big\} \ge r^{-1}\ln(n)n^{-1}(1+o_n(1)). \tag{8.2.27}$$

The right sides of (8.2.26) and (8.2.27) coincide up to a factor $1+o_n(1)$, and this allows us to say that the lower bound (8.2.27) is sharp and that the series estimator $\bar g^X(\lambda, J_n')$ is asymptotically minimax (optimal, efficient). Of course, J_n' depends on parameter r of the analytic class, and this is why the E-estimator chooses a cutoff using data, or we may say that E-estimator is an adaptive estimator because it adapts to an underlying class of spectral densities.

There is another interesting conclusion from (8.2.23) that can be made about a reasonable estimator that does not require adaptation to the class (8.1.8) and may choose a cutoff J_n a priori before getting data. Let us consider $J_n = J_n^*$ which is the largest integer smaller than $(c_n/2)\ln(n)$ where $c_n \to \infty$ as slow as desired, say $c_n = \ln(\ln(\ln(n)))$. Then a direct calculation shows that

$$\frac{\mathrm{MISE}(\bar g^X(\lambda, J_n^*), g^X(\lambda))}{\int_{-\pi}^{\pi}[g^X(\lambda)]^2 d\lambda} = c_n\ln(n)n^{-1}(1+o_n(1)). \tag{8.2.28}$$

Note that the rate of the MISE convergence is just "slightly" slower than the minimax rate $\ln(n)n^{-1}$.

As we have seen, estimation of spectral densities is an exciting topic with rich history, fascinating asymptotic theory, and numerous practical applications.

8.3 Bernoulli Missing

Similarly to the previous section, in what follows $X_1, \dots, X_n$ is a realization of a zero-mean and second-order stationary time series. Additionally, only to simplify notation, it is also assumed that $\mathbb{P}(X_t = 0) = 0$. The new here is that some observations in the realization may be missed. The missing mechanism is described by a stationary Bernoulli time series $\{A_t\}$ which is independent of $\{X_t\}$ and such that A_t takes on either value 0 or 1. To avoid a case when all observations are missed, it is assumed that $\mathbb{P}(A_t = 1) > 0$. Recall that in Chapters 4 and 5 availability variables $A_1, A_2, \dots$ were independent Bernoulli variables; in

this section they are depended and this is why we refer to them as a Bernoulli time series. Then, similarly to our previous missing models, the available observations are

$$Y_l := A_l X_l, \quad l = 1, 2, \dots, n. \tag{8.3.1}$$

As a result, X_l is missed whenever $A_l = 0$. The above-made assumption $\mathbb{P}(X_t = 0) = 0$ implies that $\mathbb{P}(A_l = I(Y_l \neq 0)) = 1$, and hence we do not need to keep track of the time series $\{A_t\}$.

The problem is to estimate the spectral density $g^X(\lambda)$ of a hidden underlying process $\{X_t\}$ given observations (8.3.1). In this section two models of the missing mechanism are considered, and more models will be presented in the next section.

The first missing model is called a *Markov–Bernoulli* model and it is generated by a Markov–Bernoulli time series (chain) $\{A_t\}$. Let us recall some basic facts about Markov chains. A discrete time Markov chain is a time series (stochastic process) that undergoes transitions from one state to another, between a finite or countable number of possible states. In our special case of the Bernoulli process there are just two states: 0 and 1. This random process is memoryless meaning that the probability of transition into the next state depends only on the current state and not on the sequence of events that preceded it. This specific kind of "memorylessness" is called the Markov property. The changes of state of the Markov chain are called transitions, and the probabilities associated with various state-changes are called transition probabilities. If transition probabilities do not change in time, then the Markov chain is called stationary. An example is the dietary habits of a creature who eats either grapes or lettuce exactly once a day conforming to the following habits. If it ate grapes today then tomorrow, regardless of what it ate in previous days, it will eat grape with probability 1/3 and lettuce with probability 2/3; do you see that these two probabilities are correctly added to 1? If it ate lettuce today then tomorrow, regardless of what it ate in previous days, it will eat grape with probability 3/5 and lettuce with probability 2/5. Suppose that an experiment on the creature (say a blood test or a training session) is conducted only on the day when grapes are eaten. Then an observation is missed at the day when lettuce is eaten. Correspondingly, the transition probabilities for the availability are $\mathbb{P}(A_t = 1|A_{t-1} = 1) = 1/3$, $\mathbb{P}(A_t = 0|A_{t-1} = 1) = 2/3$, $\mathbb{P}(A_t = 1|A_{t-1} = 0) = 3/5$, and $\mathbb{P}(A_t = 0|A_{t-1} = 0) = 2/5$. This is an example of a Markov–Bernoulli missing.

For a stationary Markov–Bernoulli missing mechanism, it is of a special interest to learn about random length of the batch of missing observations. Denote this length as L and, as an example, consider a time series with $\dots, A_t = 1, A_{t+1} = 0, A_{t+2} = 0, \dots, A_{t+L} = 0, A_{t+L+1} = 1, \dots$ Here we observe, after the available observation X_t, the batch of missing observations $X_{t+1}, \dots, X_{t+L}$ of length L. Given that $L \geq 1$, the distribution of L is geometric with $\mathbb{P}(L = k) = \alpha^{k-1}(1 - \alpha)$, $k = 1, 2, 3, \dots$ where $\alpha = \mathbb{P}(A_{t+1} = 0|A_t = 0)$. The mean length $\mathbb{E}\{L\}$ of the batch is $(1-\alpha)\alpha^{-1}$ and the variance is $(1-\alpha)\alpha^{-2}$. Another interesting property, which shows that the probability of larger batches of missing observations decreases exponentially, is described by the inequality

$$\mathbb{P}(L > k) \leq [|\ln(\alpha)|^{-1}(1 - \alpha)]e^{-k|\ln(\alpha)|}. \tag{8.3.2}$$

To conclude our brief introduction to Markov chains, let us note that a Markov chain of order m (or a Markov chain with memory m), where m is a finite positive integer, is a process satisfying

$$\mathbb{P}(A_t = a_t|A_{t-1} = a_{t-1}, A_{t-2} = a_{t-2}, \dots)$$
$$= \mathbb{P}(A_t = a_t|A_{t-1} = a_{t-1}, A_{t-2} = a_{t-2}, \dots, A_{t-m} = a_{t-m}).$$

In other words, only last m past states define the future state. Note that the classical Markov chain may be referred to as the Markov chain of order 1.

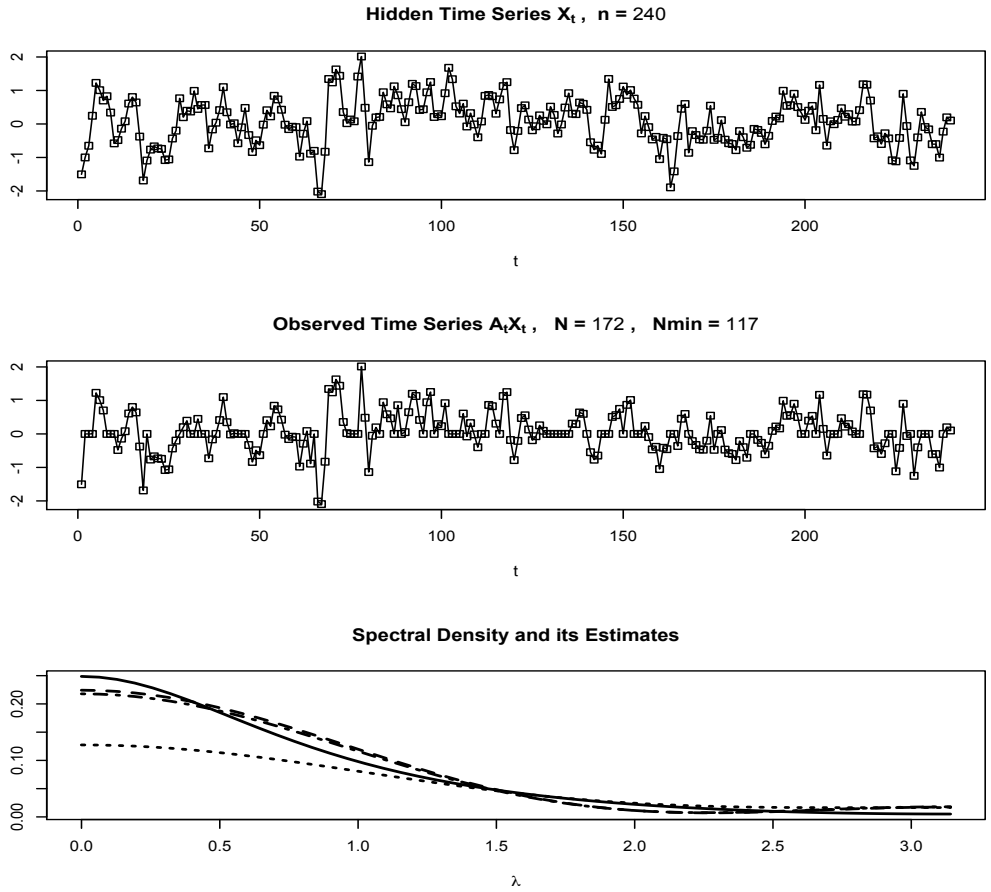

Figure 8.2 *Markov–Bernoulli missing mechanism. Markov chain is generated according to transition probabilities* $\mathbb{P}(A_{t+1} = 0|A_t = 0) = \alpha$ *and* $\mathbb{P}(A_{t+1} = 1|A_t = 1) = \beta$. *The top diagram shows* n *realizations of an underlying Gaussian ARMA(1,1) time series* X_t *defined in Figure 8.1 only here parameters are* $a = 0.4$, $b = 0.5$, $\sigma = 0.5$. *The middle diagram shows the observed time series* $\{A_tX_t\}$ *with missing observations. Set* $N_j := \sum_{l=1}^{n-j} A_lA_{l+j}$, *and then* $\mathrm{N} := N_0$ *and* $\mathrm{Nmin} := \min_{0\le j\le c_{J0sp}+c_{J1sp}\ln(n)} N_j$ *are shown in the title. The bottom diagram shows by the solid line the underlying spectral density as well as the following three estimates. The proposed E-estimate is shown by the dashed line. The naïve E-estimate (the dotted line) is based on available observations of time series* $\{A_tX_t\}$ *shown in the middle diagram. Oracle's E-estimate, based on the hidden underlying realizations of* $\{X_t\}$, *is shown by the dot-dashed line. [n = 240, sigma = 0.5, a = 0.4, b = 0.5, alpha = 0.4, beta = 0.8, cJ0sp = 2, cJ1sp = 0.5, cTHsp = 4]*

Figure 8.2 helps us to understand how a Markov–Bernoulli missing mechanism performs. The top diagram shows realization of a Gaussian ARMA(1, 1) time series $\{X_t\}$. Recall that the top diagram in Figure 8.1 also shows us a realization of a Gaussian ARMA(1, 1) time series, only here we use positive values for parameters a and b (see the caption and compare with the time series in the top diagram in Figure 8.1 where negative parameters are used). The positive parameters create a slowly oscillating time series, it is even possible to imagine a deterministic periodic (seasonal) component, but here we do know that everything is purely stochastic. It is natural to expect that in the spectral domain low frequencies dominate high frequencies, and the solid line in the bottom diagram, which is the plot of the underlying spectral density, supports this conclusion. The observed time series $\{A_tX_t\}$ is shown in the

middle diagram. Note that the missing data are easily visualized, and hence there is no need to plot the time series $A_1, \ldots, A_n$. Another observation, based on the middle diagram, is that some batches of missing observations are relatively long. The number $N := \sum_{l=1}^{n} A_l$ of available observations is shown in the title of the middle diagram. The particular number $N = 172$ shows that almost 30% of observations are missed. Can a data-driven spectral density estimator be proposed that performs reasonably well under these circumstances? Further, is the number N the only one to look after? These are the questions that we would like to explore.

Let us explain how a data-driven E-estimator, based on missing data (8.3.1), can be constructed. The aim is to propose a "good" estimate of the autocovariance function $\gamma^X(j)$ of the time series of interest $\{X_t\}$. Our approach is as follows. We begin with a sample mean autocovariance estimator of the observed time series $\{Y_t\} := \{A_t X_t\}$, and then decide on what changes should be done to estimate $\gamma^X(j)$. Using independence between time series $\{A_t\}$ and $\{X_t\}$, together with the assumption that $\mathbb{E}\{X_t\} = 0$, we conclude that $\mathbb{E}\{A_t X_t\} = 0$. Then the sample mean autocovariance is

$$\hat{\gamma}^Y(j) := (n-j)^{-1} \sum_{l=1}^{n-j} Y_l Y_{l+j} = (n-j)^{-1} \sum_{l=1}^{n-j} (A_l X_l)(A_{l+j} X_{l+j}). \tag{8.3.3}$$

Using the assumed second-order stationarity of $\{X_t\}$ and stationarity of $\{A_t\}$, we can evaluate the expectation of the sample mean autocovariance. Write,

$$\begin{aligned} \mathbb{E}\{\hat{\gamma}^Y(j)\} &= \mathbb{E}\{(n-j)^{-1} \sum_{l=1}^{n-j} (A_l X_l)(A_{l+j} X_{l+j})\} \\ &= (n-j)^{-1} \sum_{l=1}^{n-j} \mathbb{E}\{A_l A_{l+j}\} \mathbb{E}\{X_l X_{l+j}\} \\ &= \gamma^X(j)[(n-j)^{-1} \sum_{l=1}^{n-j} \mathbb{E}\{A_l A_{l+j}\}] = \gamma^X(j) \mathbb{E}\{A_1 A_{1+j}\}. \end{aligned} \tag{8.3.4}$$

Note that the expectation $\mathbb{E}\{A_1 A_{1+j}\}$, which we see in (8.3.4), looks like the autocovariance but it is not because the expectation of A_t is not zero.

We conclude that the expectation of the estimator $\hat{\gamma}^Y(j)$, based on the observed process Y_t, is the product of the underlying autocovariance of interest $\gamma^X(j)$ and the function $\mathbb{E}\{A_1 A_{1+j}\} = \mathbb{P}(A_1 A_{1+j} = 1)$. The function $\mathbb{E}\{A_1 A_{1+j}\}$ can be estimated by its sample mean, and this yields the following plug-in sample mean estimator of the autocovariance of interest $\gamma^X(j)$,

$$\hat{\gamma}^X(j) := \frac{(n-j)^{-1} \sum_{l=1}^{n-j} Y_l Y_{l+j}}{(n-j)^{-1} \sum_{l=1}^{n-j} I(Y_l Y_{l+j} \neq 0)} = \frac{\sum_{l=1}^{n-j} Y_l Y_{l+j}}{\sum_{l=1}^{n-j} I(Y_l Y_{l+j} \neq 0)}. \tag{8.3.5}$$

Similarly to Section 4.1, we need to comment on the term $N_j := \sum_{l=1}^{n-j} I(Y_l Y_{l+j} \neq 0) = \sum_{l=1}^{n-j} A_l A_{l+j}$ which is used in the denominator of (8.3.5). Theoretically the number of available pairs of observations may be zero, and then using the assumed $0/0 := 0$ in (8.3.5) we get $\hat{\gamma}^X(j) = 0$. This is a reasonable outcome for the case when no information about $\gamma^X(j)$ is available. Another remedy is to consider only samples with $N_{\min} := \min_{\{0 \le j \le c_{J0sp} + c_{J1sp} \ln(n)\}} N_j > k$ for some $k \ge 0$, because $c_{J0sp} + c_{J1sp} \ln(n)$ is the largest frequency used by the E-estimator (recall (2.2.4)).

Is the plug-in estimator (8.3.5) unbiased whenever $N_j > k \ge 0$? It is often not the case when a statistic is plugged in the denominator of an unbiased estimator, and hence it is of interest to check this property. Using independence between $\{A_t\}$ and $\{X_t\}$ we can write,

$$\mathbb{E}\{\hat\gamma^X(j)|N_j>k\} = \mathbb{E}\Big\{\mathbb{E}\Big\{\frac{\sum_{l=1}^{n-j}(A_lX_l)(A_{l+j}X_{l+j})}{\sum_{l=1}^{n-j}A_lA_{l+j}}\Big|A_1,\dots,A_n,N_j>k\Big\}\Big\}$$

$$= \mathbb{E}\Big\{\frac{\sum_{l=1}^{n-j}A_lA_{l+j}\mathbb{E}\{X_lX_{l+j}\}}{\sum_{l=1}^{n-j}A_lA_{l+j}}\Big|N_j>k\Big\} = \gamma^X(j). \tag{8.3.6}$$

This establishes that, whenever $N_{\min} > 0$, the proposed autocovariance estimator (8.3.5) is unbiased, and it can be used to construct the spectral density E-estimator $\hat g^X(\lambda)$.

Now let us return to Figure 8.2. The title of the middle diagram indicates $N_{\min} = 117$, and this tells us about complexity of the problem of the spectral density estimation with missing data when the size $n = 240$ of the hidden time series is decreased to $N = 172$ available observations, and then the minimal number of available pairs for estimation of the autocovariance is decreased to 117, that is more than the twofold decrease of n. It is also of interest to present the underlying $N_0,\dots,N_5$ that are 172, 134, 120, 117, 122 and 119, respectively. We may conclude that, in the analysis of a time series with missing observations, it is important to take into account not only n and N but also $N_{\min}$.

The bottom diagram in Figure 8.2 exhibits three estimates and the underlying spectral density of interest $g^X(\lambda)$ (the solid line). The dashed line is the proposed data-driven E-estimate. Note that it correctly exhibits the slowly decreasing shape of the underlying density (compare with the solid line which is the underlying spectral density). The dotted line shows us the E-estimate based on $\hat\gamma^Y(j)$, in other words, this is a naïve estimate which ignores the missing data and deals with $\{Y_t\}$ like it is the time series of interest. While overall it is a poor estimate of the underlying spectral density, note that it correctly shows smaller power of the observed time series at low frequencies. The dot-dashed line shows us oracle's E-estimate based on the hidden time series $\{X_t\}$ shown in the top diagram. The bottom diagram allows us to compare performance of the same E-estimator based on three different datasets, and based on this single simulation we may conclude that the proposed data-driven E-estimator performs relatively well, and ignoring the missing, as the naïve E-estimator does, is a mistake. The interested reader is advised to repeat this figure, possibly with different parameters, to get first-hand experience in dealing with the Markov-Bernoulli missing mechanism.

Now let us consider a different mechanism of creating missing observations. Here the number L of consecutive missing observations, that is the length of a batch of missing observations, is generated by a random variable. Correspondingly, we will generate missing observations by choosing the distribution of L and refer to this missing mechanism as batch-Bernoulli. The following example clarifies the definition. Suppose that each hour we need to conduct an experiment whose outcomes create a time series of hourly observations. However, this task is of lower priority with respect to other jobs that may arrive in batches with each job requiring one hour to be fulfilled. As a result we may have intervals of time when the experiment is not conducted. A modification of that example is when the experiment cannot be performed if the equipment malfunctions and then a random number of hours is required to fix the equipment.

A distribution of L, which is often used to model batches, is Poisson with $\mathbb{E}\{L\} = \lambda$, and definition of this distribution can be found in Section 1.3. To simulate the corresponding batch-Bernoulli missing, we generate a sample $L_1, L_2, \dots$ of independent random variables from L. If $L_1 = 0$ then no missing occurs and if $L_1 = k > 0$ then k consecutive observations $X_1,\dots,X_k$ are missed, etc. The following inequality sheds light on how large Poisson batches can be,

$$\mathbb{P}(L \ge k) \le e^{-\lambda}k^{-k}(e\lambda)^k = e^{-k\ln(k/e\lambda)-\lambda}, \ \ k > \lambda. \tag{8.3.7}$$

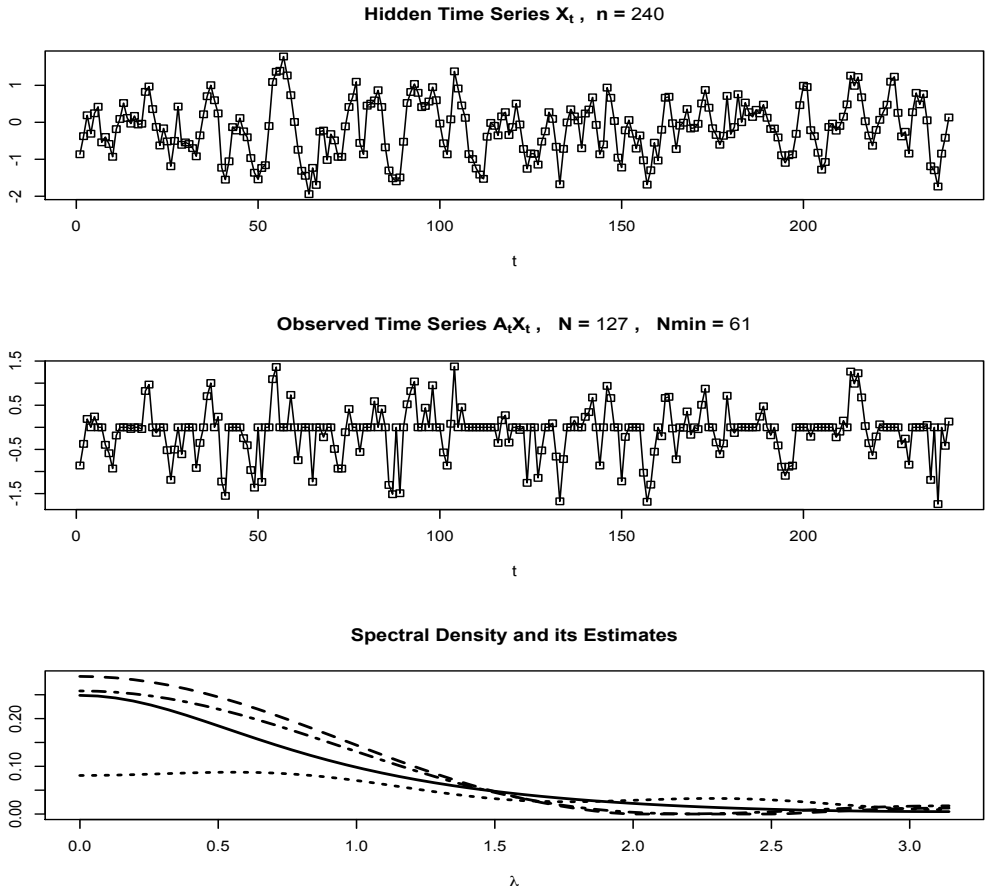

Figure 8.3 *Batch-Bernoulli missing mechanism. Lengths $L_1, L_2, \ldots$ of missing batches are independent Poisson random variables with mean $\mathbb{E}\{L\} = \lambda$. Otherwise the simulation and structure of diagrams are similar to Figure 8.2.* {*Parameter λ is controlled by the argument lambda.*} *[n = 240, sigma = 0.5, a = 0.4, b = 0.5, lambda=0.5, cJ0sp = 2, cJ1sp = 0.5, cTHsp = 4]*

For a batch-Bernoulli missing the relation (8.3.6) still holds and hence the same spectral density E-estimator may be used.

Figure 8.3 illustrates the batch-Bernoulli missing mechanism, and here $\lambda = 0.5$. Apart of the new missing mechanism, the simulation and diagrams are the same as in Figure 8.2. The top diagram shows a realization of the ARMA process. The middle diagram shows us the missing pattern, and it sheds new light on the name of the missing mechanism. We can also note from the title that in this particular simulation, the Poisson batches decreased the number $n = 240$ of hidden observations to just $N = 127$ available observations, in other words, almost a half of hidden underlying observations is missed. Further, the minimal number of available pairs for calculation autocovariance coefficients of the E-estimator is $N_{\min} = 61$, and it is almost a quarter of $n = 240$. Let us also add the information about the underlying $N_0, \ldots, N_5$ that are 127, 71, 66, 62, 61 and 64, respectively. Not surprisingly, these small numbers of pairs, available for calculation of the autocovariance, take a toll on the E-estimate (compare the dashed line of the E-estimate with the dot-dashed line of the oracle's E-estimate based on $n = 240$ hidden realizations of $\{X_t\}$). At the same time, the proposed estimate still nicely exhibits shape of the underlying spectral density and it is

dramatically better than the naïve E-estimate (the dotted line) which ignores the missing and treats $\{Y_t\}$ as the time series of interest. It is a good exercise to repeat this figure with different parameters and then analyze patterns in time series and performance of estimators.

8.4 Amplitude-Modulated Missing

The considered problem is to estimate the spectral density $g^X(\lambda)$ of a time series $\{X_t\}$. Observations of the time series are hidden and instead observations of a time series $\{Y_t\} := \{U_tA_tX_t\}$ are available. Here, similarly to the previous section, $\{A_t\}$ is a stationary Bernoulli time series which defines an underlying missing mechanism, $\{U_t\}$ is a zero-mean and second-order stationary time series which performs amplitude-modulation. Let us make several additional assumptions. Similarly to the previous section we assume that $\mathbb{P}(X_t = 0) = 0$. To separate the effects of A_t and U_t, it is assumed that $\mathbb{P}(U_t = 0) = 0$ and this prevents U_t from affecting the missing and allows us to say that $\{U_t\}$ is a pure amplitude-modulating time series. (Of course, a time series $\{A_t\}$ may be also looked at as an amplitude-modulating time series, and this terminology is well accepted in the literature.) Further, it is assumed that $U_1, U_2, \ldots$ are independent and identically distributed realizations of a random variable U such that

$$\mathbb{P}(U = 0) = 0, \quad \mathbb{E}\{U\} = \mu \neq 0, \quad \mathbb{E}\{U^2\} = \mu_2. \tag{8.4.1}$$

Further, it is assumed that the times series $\{U_t\}$, $\{A_t\}$ and $\{X_t\}$ are mutually independent.

Is it possible to estimate the spectral density $g^X(\lambda)$ of an underlying time series $\{X_t\}$ based on amplitude-modulated observations $Y_t = U_tA_tX_t$, $t = 1, 2, \ldots, n$? To answer this question, let us look at the autocovariance function $\gamma^Y(j)$ of the observed time series $\{Y_t\}$. Introduce a sample mean autocovariance,

$$\hat{\gamma}^Y(j) := (n-j)^{-1}\sum_{l=1}^{n} Y_l = (n-j)^{-1}\sum_{l=1}^{n} A_lU_lX_l. \tag{8.4.2}$$

Because $\{X_t\}$ is zero-mean and it is assumed that the three time series are mutually independent, the observed time series $\{Y_t\}$ is also zero-mean. Then, using (8.4.1) we may write,

$$\begin{aligned}\mathbb{E}\{\hat{\gamma}^Y(j)\} &= \mathbb{E}\{(U_1A_1X_1)(U_{1+j}A_{1+j}X_{1+j})\}\\ &= \mathbb{E}\{U_1U_{1+j}\}\mathbb{E}\{A_1A_{1+j}\}\mathbb{E}\{X_1X_{1+j}\}\\ &= [\mu_2 I(j=0) + \mu^2 I(j\neq 0)]\mathbb{E}\{A_1A_{1+j}\}\gamma^X(j).\end{aligned} \tag{8.4.3}$$

Equation (8.4.3) sheds light on a possibility to solve the problem of estimation of the spectral density $g^X(\lambda)$. Namely, if parameters μ_2 and μ are known, then it is possible to estimate the spectral density using the approach of the previous section. Further, if the two parameters are unknown but a sample from U is available, then the two parameters can be estimated by sample mean estimators. Note that in both cases some extra information is needed. On the other hand, (8.4.3) indicates that, based solely on amplitude-modulated data we cannot consistently estimate the spectral density $g^X(\lambda)$. In other words, the considered modification is destructive.

Despite the above-made gloom conclusion, the following approach may be a feasible remedy in some practical situations. Introduce a function $s^X(\lambda)$ which is called the spectral *shape* or the shape of spectral density,

$$s^X(\lambda) := \pi^{-1}\sum_{j=1}^{\infty}\gamma^X(j)\cos(j\lambda). \tag{8.4.4}$$

Note that $g^X(\lambda) = (2\pi)^{-1}\gamma^X(0) + s^X(\lambda)$, and hence in a graphic the spectral shape is just shifted (in vertical direction) spectral density, and the shift is such that the integral of the shape over $[0, \pi]$ is zero. Further, apart of the variance of X_t, the spectral shape provides us with all values of the autocovariance function $\gamma^X(j)$, $j > 0$. Further, in practical applications the spectral density is often of interest in terms of its modes, and then knowing either the spectral density, or the spectral shape, or the scaled spectral shape $s^X(\lambda, \mu)$ defined as

$$s^X(\lambda, \mu) = \mu^2 s^X(\lambda) =: \pi^{-1}\sum_{j=1}^{\infty}(\mu^2\gamma^X(j))\cos(j\lambda) =: \pi^{-1}\sum_{j=1}^{\infty}\eta^X(j)\cos(j\lambda) \tag{8.4.5}$$

is equivalent. Recall that $\mu := \mathbb{E}\{U\}$ is an unknown parameter (the mean) of the amplitude-modulating distribution.

The above-made remark makes the problem of estimation of the scaled spectral shape of a practical interest. Furthermore, estimation of the scaled shape of spectral density is possible based on the available amplitude-modulated time series $\{Y_t\}$. Indeed, consider the estimator

$$\hat{\eta}^X(j) := \frac{\sum_{l=1}^{n-j} Y_l Y_{l+j}}{\sum_{l=1}^{n-j} I(Y_l Y_{l+j} \neq 0)}, \ \ j \geq 1 \tag{8.4.6}$$

of coefficients $\eta^X(j)$ in the cosine expansion (8.4.5). Following Section 8.3, set $N_j := \sum_{l=1}^{n-j} A_l A_{l+j}$, $N := N_0$ and $N_{\min} := \min_{\{0 \leq j \leq c_{J0sp} + c_{J1sp}\ln(n)\}} N_j$. Let us explore the expectation of the estimator (8.4.6) given $N_j > 0$. Write,

$$\begin{aligned}
\mathbb{E}\{\hat{\eta}^X(j)|N_j > 0\} &= \mathbb{E}\Big\{\mathbb{E}\Big\{\frac{\sum_{l=1}^{n-j} Y_l Y_{l+j}}{\sum_{l=1}^{n-j} I(Y_l Y_{l+j} \neq 0)}|A_1, \ldots, A_n, N_j > 0\Big\}\Big\} \\
&= \mathbb{E}\Big\{\frac{\sum_{l=1}^{n-j}(A_l A_{l+j})\mathbb{E}\{U_l U_{l+j}\}\mathbb{E}\{X_l X_{l+j}\}}{\sum_{l=1}^{n-j} A_l A_{l+j}}\Big|N_j > 0\Big\} \\
&= \mu^2\gamma^X(j) = \eta_j^X, \quad j \geq 1.
\end{aligned} \tag{8.4.7}$$

We conclude that whenever $N_{\min} > 0$ the plug-in sample mean estimator (8.4.6) is unbiased for all j used by the E-estimator. Hence, this estimator can be used for construction of the E-estimator $\hat{s}^X(\lambda, \mu)$ of the scaled shape. Note that the parameter μ is still unknown, and hence the underlying spectral shape $s^X(\lambda)$ will be estimated up to an unknown factor $\mu^2 = [\mathbb{E}\{U\}]^2$.

Figure 8.4 illustrates considered problem of amplitude-modulated time series. The top diagram shows a realization of a hidden time series of interest. The observed amplitude-modulated time series is shown in the middle diagram. The used Markov-Bernoulli process $\{A_t\}$ is the same as in Figure 8.2. The process $U_t := \mu U_t^*/\mathbb{E}\{U_t^*\}$ where $U_1^*, \ldots, U_n^*$ are independent random variables generated according to one of our corner densities; here the density is the Normal. The title shows that only $N = 180$ from the underlying $n = 240$ observations are available. This is not too bad, but $N_{\min} = 125$ informs us about a heavy missing. Let us also present information about $N_0, \ldots, N_5$ that are 180, 143, 139, 130, 128 and 125, respectively. We see that the missing implies an almost twofold decrease in the number of available pairs for calculating the autocovatiance function. In the bottom diagram the underlying scaled shape $s^X(\lambda, \mu)$ is shown by the solid line, the E-estimate and the oracle's estimate, based on the underlying realization of $\{X_t\}$, are shown by the dashed and dot-dashed lines, respectively, the dotted curve shows the naïve E-estimate based on the available observations $Y_1, \ldots, Y_n$. The E-estimate is close to its oracle's benchmark while the naïve estimate, which ignores the amplitude-modulated nature of the observed time series, is far from being even close to the underlying scaled shape.

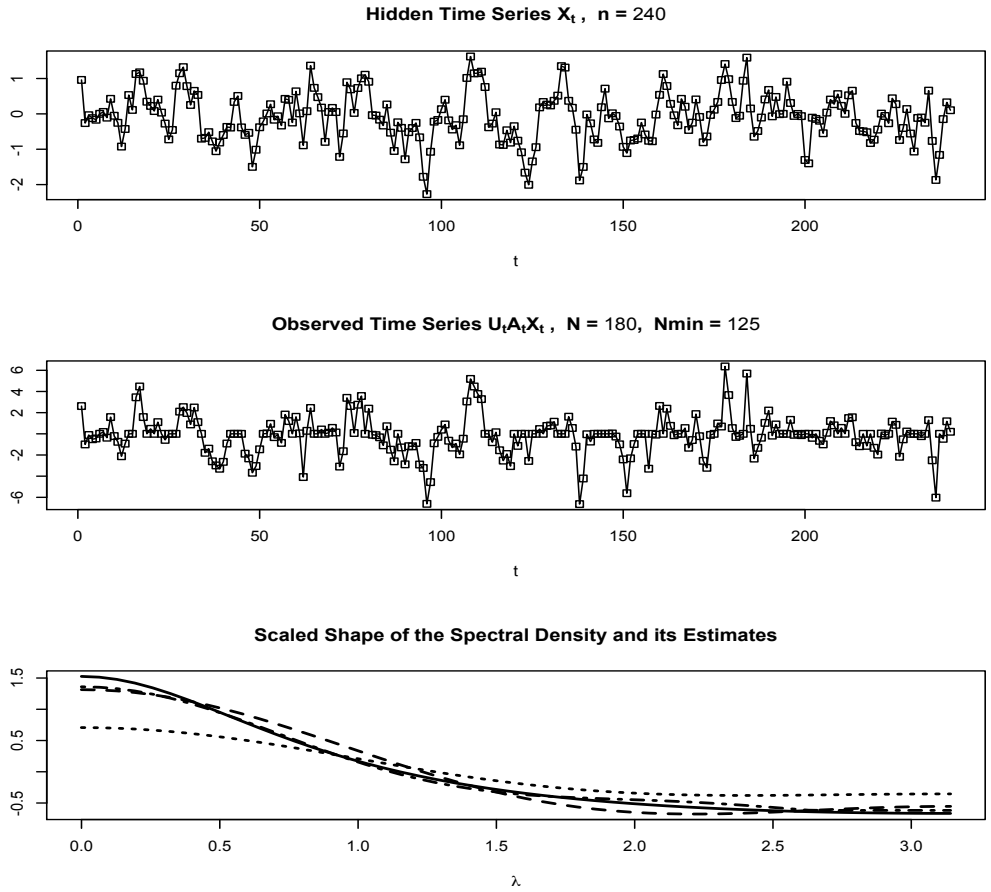

Figure 8.4 *Amplitude-modulated time series and estimation of the scaled shape of the spectral density. Amplitude-modulated observations* $Y_t := U_t A_t X_t$, $t = 1, 2, \ldots, n$ *are generated by a Markov-Bernoulli time series* $\{A_t\}$ *used in Figure 8.2, and* $U_1, U_2, \ldots, U_n$ *is a sample from a scaled corner density whose mean* $\mathbb{E}\{U_t\} = \mu$. *The structure of diagrams is identical to Figure 8.2. For the considered setting only estimation of the scaled shape* $s^X(\lambda, \mu)$ *of the spectral density is possible, and the scaled shape and its estimates are shown in the bottom diagram.* {*Parameter* μ *is controlled by the argument mu, the choice of corner function is controlled by argument corn.*} *[n = 240, sigma = 0.5, a = 0.4, b = 0.5, alpha = 0.4, beta = 0.8, corn = 2, mu = 3, cJ0sp = 2, cJ1sp = 0.5, cTHsp = 4]*

Now let us consider another interesting example of an amplitude-modulated time series. Assume that $Z_1, \ldots, Z_n$ are independent realizations of (sample from) a Poisson random variable Z with the mean λ, and define the observed time series as $\{Y_t\} := \{Z_t X_t\}$. In this example, variables Z_t create both the missing (when $Z_t = 0$) and the amplitude-modulation when $Z_t > 0$. The example is of interest because it happens in practical applications and because, as we will see shortly, the Poisson amplitude-modulation is not destructive.

Let us explain a proposed E-estimation of the spectral density $g^X(\lambda)$. Note that $\mathbb{E}\{I(Z_t = 0)\} = \mathbb{P}(Z_t = 0) = e^{-\lambda}$, and hence the parameter λ can be estimated by the plug-in sample mean estimator

$$\hat{\lambda} := -\ln(\sum_{l=1}^{n} I(Y_l = 0)/n). \tag{8.4.8}$$

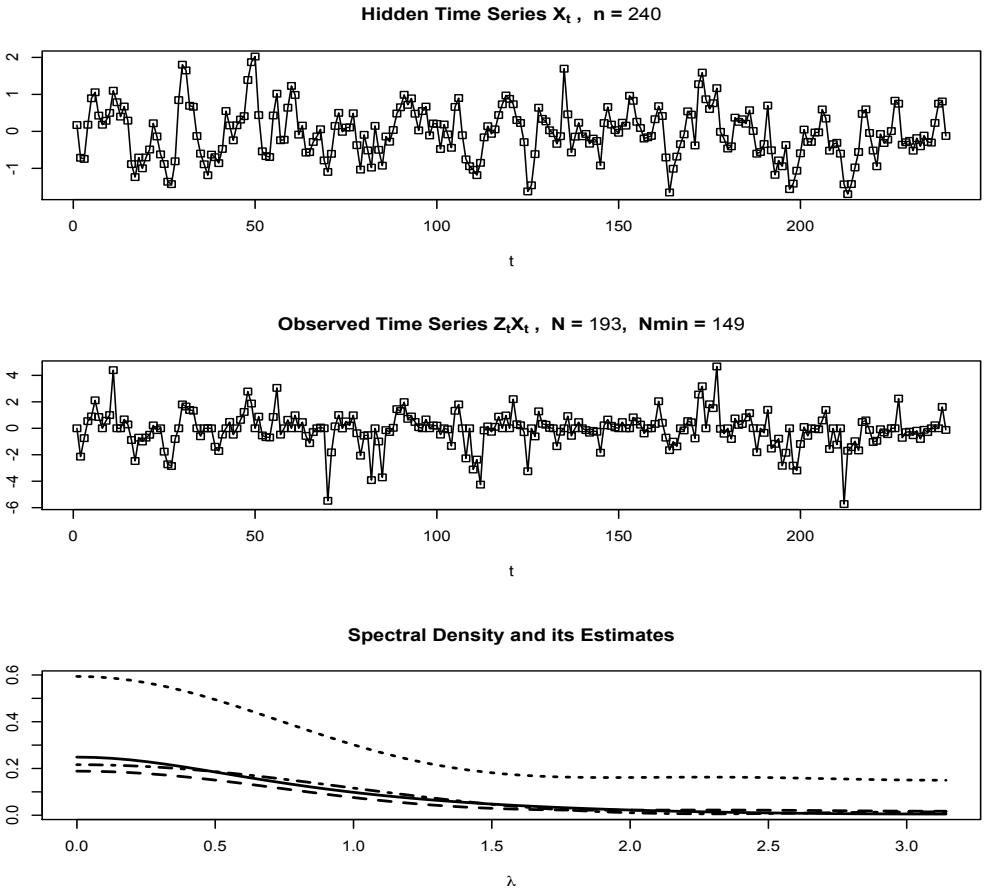

Figure 8.5 *Amplitude-modulation by Poisson variable. Observed time series is* $Y_t = Z_tX_t$, $t = 1, 2, \ldots, n$ *where* $\{X_t\}$ *is the ARMA process defined in the caption of Figure 8.2 and* $Z_1, Z_2, \ldots, Z_n$ *is a sample from a Poisson random variable* Z *with the mean* $\mathbb{E}\{Z\} = \lambda$. *The structure of diagrams is similar to those in Figure 8.2.* {*The argument lambda controls* λ.} *[n = 240, sigma = 0.5, a = 0.4, b = 0.5, lambda = 1.5, cJ0sp = 2, cJ1sp = 0.5, cTHsp = 4]*

Further, we can write

$$\gamma^Y(j) = [(\lambda + \lambda^2)I(j = 0) + \lambda^2 I(j \neq 0)]\gamma^X(j), \;\; j = 0, 1, \ldots \tag{8.4.9}$$

This allows us to propose an estimator of the autocovariance function

$$\hat{\gamma}^X(j) := [(\hat{\lambda} + \hat{\lambda}^2)I(j = 0) + \hat{\lambda}^2 I(j \neq 0)]^{-1}\hat{\gamma}^Y(j), \;\; j = 0, 1, \ldots \tag{8.4.10}$$

In its turn, the autocovariance estimator yields the E-estimator of the spectral density.

Figure 8.5, whose structure is similar to Figure 8.2, illustrates both the setting and the E-estimator. The top diagram shows the underlying time series of interest. The middle diagram shows the observed amplitude-modulated time series, as well as the number $N := \sum_{l=1}^{n} I(Z_l X_l \neq 0) = 193$ of available observations, and $N_{\min} = 149$. Also, let us present numbers $N_0, \ldots, N_5$ of available pairs 193, 151, 157, 150, 149 and 153, respectively. Returning to the middle diagram, please note how visually different are realizations of the amplitude-modulated time series and the underlying one. Further, note the difference in

their scales. It is difficult to believe that it is possible to restore the underlying spectral density of the hidden time series $\{X_t\}$ after its amplitude-modulation by the Poisson variable. Nonetheless, the bottom diagram shows that the spectral density E-estimate is close to the oracle's estimate based on the hidden time series, and overall it is very good. On the other hand, the naïve spectral density E-estimate of Section 8.2, based on observations $Z_1X_1, \ldots, Z_nX_n$, is clearly poor. Note that it indicates a dramatically larger overall power of the observed time series, especially on low frequencies.

The overall conclusion is that it is absolutely prudent to pay attention to a possible modification of an underlying time series of interest.

8.5 Censored Time Series

The aim is to explore the problem of estimation of the spectral density for a time series modified by a right censoring. An underlying zero-mean and stationary time series of interest $\{X_t\}$ is hidden and not available for a statistical inference. Instead, a bivariate time series $\{V_t, \Delta_t\}$ is observed where $V_t := \min(X_t, C_t)$, $\Delta_t := I(X_t \le C_t)$, and $\{C_t\}$ is a stationary censoring time series. It what follows it is assumed that all random variables are continuous, their distributions are unknown, time series $\{X_t\}$ and $\{C_t\}$ are independent, and $\beta_{X_t} \le \beta_{C_t}$ (recall that β_Z denotes the upper bound of the support of Z). Recall that the right censoring (RC) was discussed in great details in Chapter 6, only here we no longer assume that the variables are nonnegative because we are dealing with zero-mean time series.

To explore the problem of estimation of the spectral density $g^X(\lambda)$, we begin with a formula for a joint density of two pairs of observed variables,

$$f^{V_1,V_{1+j},\Delta_1,\Delta_{1+j}}(v_1, v_{1+j}, 1, 1) = f^{X_1,X_{1+j}}(v_1, v_{1+j})G^{C_1,C_{1+j}}(v_1, v_{1+j}), \tag{8.5.1}$$

where $G^{C_1,C_{1+j}}(v_1, v_{1+j}) = \mathbb{P}(C_1 > v_1, C_{1+j} > v_{1+j})$ is the bivariate survival function. This formula allows us to write,

$$\mathbb{E}\{V_1V_{1+j}\Delta_1\Delta_{1+j}\}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_1v_{1+j}f^{X_1,X_{1+j}}(v_1, v_{1+j})G^{C_1,C_{1+j}}(v_1, v_{1+j})dv_1dv_{1+j}. \tag{8.5.2}$$

This relation points upon a possible solution. Recall that $\{X_t\}$ is a zero-mean and stationary time series. Using this assumption, together with (8.5.1), allows us to write down the autocovariance function of interest as

$$\gamma^X(j) := \mathbb{E}\{X_1X_{1+j}\}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_1v_{1+j}f^{X_1,X_{1+j}}(v_1, v_{1+j})dv_1dv_{1+j}$$
$$= \mathbb{E}\Big\{\frac{V_1V_{1+j}\Delta_1\Delta_{1+j}}{G^{C_1,C_{1+j}}(V_1, V_{1+j})}\Big\}. \tag{8.5.3}$$

In its turn, (8.5.3) implies that if the bivariate survival function $G^{C_1,C_{1+j}}(v, u)$ of the censoring time series is known (for instance from previous or extra studies), then the sample mean estimator of the autocovariance function $\gamma^X(j)$ is

$$\tilde{\gamma}^X(j) := (n-j)^{-1}\sum_{l=1}^{n-j}\frac{V_lV_{j+j}\Delta_l\Delta_{l+j}}{G^{C_1,C_{1+j}}(V_l, V_{l+j})}. \tag{8.5.4}$$

Note that the estimator is unbiased.

If the bivariate survival function is unknown, then we need to estimate it. Here we restrict our attention to the case when $C_1, C_2, \ldots, C_n$ are independent and identically distributed realizations of a random variable C. To propose an estimator, we note that

$$G^{V_t}(v) = G^{X_t}(v)G^C(v), \tag{8.5.5}$$

where $G^Z(z) := \mathbb{P}(Z > z)$ denotes the survival function of Z. This allows us to write

$$f^{V_t,\Delta_t}(v,0) = f^C(v)G^{X_t}(v) = \frac{f^C(v)G^{V_t}(v)}{G^C(v)}. \tag{8.5.6}$$

Next, recall that the hazard rate function of the censoring variable is defined as

$$h^C(v) := \frac{f^C(v)}{G^C(v)}. \tag{8.5.7}$$

Using this relation in (8.5.6) yields that

$$f^{V_t,\Delta_t}(v,0) = h^C(v)G^{V_t}(v). \tag{8.5.8}$$

In its turn, (8.5.8) implies that

$$h^C(v) = \frac{f^{V_t,\Delta_t}(v,0)}{G^{V_t}(v)} \tag{8.5.9}$$

whenever $G^{V_t}(v) > 0$.

The numerator and denominator on the right side of (8.5.9) can be consistently estimated because these are characteristics of observed random variables, then the ratio of the estimators may be used to estimate the hazard rate. Recall that the hazard rate is a characteristic of the distribution, and hence an estimator of the hazard rate can be used as a plug-in for a corresponding estimator of the survival function. There is a simple relation between a survival function and the corresponding cumulative hazard that will be used shortly.

The cumulative hazard of the censoring variable is defined as

$$H^C(v) := \int_{-\infty}^{v} h^C(y)dy. \tag{8.5.10}$$

Using (8.5.9) allows us to continue (8.5.10),

$$\begin{aligned} H^C(v) &:= \int_{-\infty}^{v} \frac{f^{V_t,\Delta_t}(y,0)}{G^{V_t}(y)}dy \\ &= \mathbb{E}\{(1-\Delta_t)I(V_t \le v)[G^{V_t}(V_t)]^{-1}\}. \end{aligned} \tag{8.5.11}$$

To use this expectation for constructing a sample mean estimator, we need to estimate the survival function of V_t. Note that the time series $\{V_t\}$ is stationary, and then the following sample mean estimator is a natural choice,

$$\hat{G}^{V_t}(v) := n^{-1}\sum_{l=1}^{n} I(V_l \ge v). \tag{8.5.12}$$

Using this estimator in (8.5.11) implies the sample mean estimator of the cumulative hazard $H^C(v)$,

$$\hat{H}^C(v) := n^{-1}\sum_{l=1}^{n} \frac{(1-\Delta_l)I(V_l \le v)}{\hat{G}^{V_l}(V_l)}. \tag{8.5.13}$$

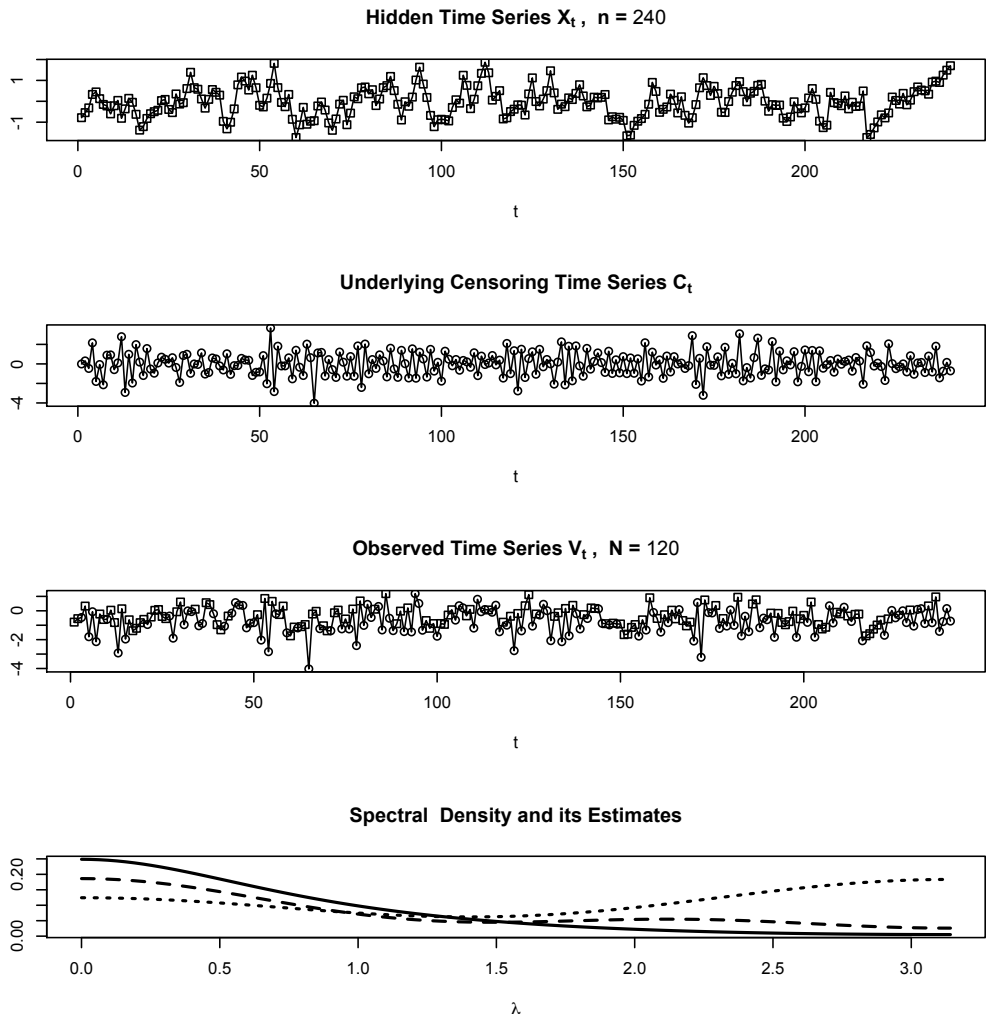

Figure 8.6 *Censored time series. The underlying time series* $\{X_t\}$ *is generated by the same Gaussian ARMA process as in Figure 8.2. The censoring time series* $\{C_t\}$ *is generated by a sample from a Gaussian variable with zero mean and standard deviation* σ_C*. In the third from the top diagram squares and circles show cases with uncensored and censored* X_t*, respectively.* $N := \sum_{l=1}^{n} \Delta_l$ *is the number of uncensored observations. In the bottom diagram the underlying spectral density* $g^X(\lambda)$*, its E-estimate and naïve E-estimate are shown by the solid, dashed and dotted lines, respectively.* {*Parameter* σ_C *of the Gaussian censoring time series is controlled by the argument sigmaC.*} *[n = 240, sigma = 0.5, a = 0.4, b = 0.5, sigmaC = 1, cJ0sp = 2, cJ1sp = 0.5, cTHsp = 4]*

As soon as the cumulative hazard is estimated, we use the familiar from Chapter 4 relation

$$G^C(v) = \exp(-H^C(v)) \tag{8.5.14}$$

to estimate the survival function of C by the plug-in estimator,

$$\hat{G}^C(v) = \exp(-\hat{H}^C(v)). \tag{8.5.15}$$

Note that in the considered case $G^{C_1,C_{1+j}}(v,u) = G^C(u)G^C(v)$, and hence estimator (8.5.15) may be used in (8.5.4) whenever the bivariate survival function is unknown.

Figure 8.6 sheds light on the setting and the proposed estimation. The top diagram shows us by squares a realization of an underlying time series $\{X_t\}$. It is generated by the same Gaussian ARMA process as in Figure 8.2. The second from the top diagram shows us by circles the censoring time series which is generated by a white Gaussian process with the unit standard deviation. Note that $\{X_t\}$ and $\{C_t\}$ have a different behavior in terms of oscillations and they have different shapes of the spectral densities. The censored time series is shown in the third from the top diagram. Here we see $V_t = \min(X_t, C_t)$, $t = 1, \ldots, n$, and if $\Delta_t = 1$ then the value of V_t is shown by the square and if $\Delta = 0$ then by the circle. In other words, uncensored observations of $\{X_t\}$ are shown by squares (and you can also see them in the top diagram) while the censored observations are hidden and instead you can see the corresponding values of $\{C_t\}$ shown by circles (they may be also observed in the second from the top diagram). Note that only $N = 120$ observations are uncensored, and this is a severe loss of information. The third diagram contains all available observations that can be used in statistical analysis. A visual analysis clearly indicates that the dynamic and spectral patterns of time series $\{V_t\}$ and $\{X_t\}$ are different. The bottom diagram shows us the underlying spectral density $g^X(\lambda)$ (the solid line) and the proposed estimate (the dashed line). The dotted line shows the spectral density E-estimate of the time series $\{V_t\}$, which is a naïve estimate that ignores the censoring.

Censored time series is a complicated stochastic modification when underlying observations are missing not at random and then substituted by values of a censoring time series. Repeated simulations, generated by Figure 8.6, may help in understanding the modification and how the proposed estimation performs.

8.6 Probability Density Estimation

Density estimation for the case of a sample $X_1, X_2, \ldots, X_n$ from a continuous random variable X, when the observations are independent, was considered in Section 2.2. In this section we would like to consider the density estimation problem for the case of a stationary time series $\{X_t\}$ of continuous variables.

Let us first briefly recall the idea of E-estimation presented in Section 2.2 for the case of a sample of independent observations. It is assumed that the density $f^X(x)$ is supported and square-integrable on $[0, 1]$, and we use the cosine basis $\varphi_0(x) := 1$, $\varphi_j(x) = 2^{1/2}\cos(\pi j x)$, $j = 1, 2, \ldots$ on $[0, 1]$. Then the density can be written as a Fourier expansion,

$$f^X(x) = \sum_{j=0}^{\infty} \theta_j \varphi_j(x), \tag{8.6.1}$$

where

$$\theta_j := \int_0^1 \varphi_j(x) f^X(x) dx \tag{8.6.2}$$

are Fourier coefficients of $f^X(x)$. The idea of E-estimation is based on the fact that the right side of (8.6.2) can be written as the expectation, namely

$$\theta_j = \mathbb{E}\{\varphi_j(X)\}. \tag{8.6.3}$$

The expectation immediately implies the following sample mean Fourier estimator,

$$\hat{\theta}_j := n^{-1} \sum_{l=1}^{n} \varphi_j(X_l). \tag{8.6.4}$$

The Fourier estimator is unbiased and

$$\mathbb{V}(\hat{\theta}_j) = n^{-1}(1 + o_j(1)). \tag{8.6.5}$$

Using this Fourier estimator by the E-estimator of Section 2.2 yields the wished density estimator. Further, Section 2.2 shows via simulations a good performance of the E-estimator for relatively small samples. In short, E-estimation is based on a good sample mean Fourier estimator.

Now let us relax the assumption about independence and consider a realization $X_1, X_2, \ldots, X_n$ of a stationary time series $\{X_t\}$. The problem is again to estimate the density $f^X(x)$, and note that now this density is marginal with respect to the joint density $f^{X_1,\ldots,X_n}(x_1, \ldots, x_n)$. Further, let us relax another assumption that the support of the density is known. Further, what will be if some of the dependent observations are missed? These are the issues that we would like to address.

First, let us relax the assumption about independence while still considering densities supported on $[0, 1]$. Let us carefully look at the Fourier estimator (8.6.4). For a stationary time series the estimator is still a sample mean estimator and unbiased. Indeed, we know that the expectation of a sum of random variables is always a sum of expectations, regardless of dependence between the random variables, and hence even if $X_1, X_2, \ldots, X_n$ are dependent we get

$$\mathbb{E}\{\hat{\theta}_j\} = \mathbb{E}\{n^{-1}\sum_{l=1}^{n}\varphi_j(X_l)\} = n^{-1}\sum_{l=1}^{n}\mathbb{E}\{\varphi_j(X_l)\}. \tag{8.6.6}$$

Using the assumed stationarity of the time series we can continue,

$$\mathbb{E}\{\hat{\theta}_j\} = n^{-1}n\int_0^1 f^X(x)\varphi_j(x)dx = \theta_j. \tag{8.6.7}$$

This proves that Fourier estimator (8.6.4) is unbiased. Further, the proof of (8.6.7) is based only on the property that the expectation $\mathbb{E}\{\varphi_j(X_t)$ does not depend on t, and this is a weaker assumption than the stationarity.

Next we need to evaluate the variance of the estimator (8.6.4). Write,

$$\mathbb{V}(\hat{\theta}_j) = \mathbb{E}\{[n^{-1}\sum_{l=1}^{n}\varphi_j(X_l) - \theta_j]^2\} = \mathbb{E}\{[n^{-1}\sum_{l=1}^{n}(\varphi_j(X_l) - \theta_j)]^2\}$$

$$= n^{-2}\sum_{l,t=1}^{n}\mathbb{E}\{(\varphi_j(X_l) - \theta_j)(\varphi_j(X_t) - \theta_j)\} = n^{-2}\sum_{l,t=1}^{n}\mathbb{E}\{\varphi_j(X_l)\varphi_j(X_t)\} - \theta_j^2. \tag{8.6.8}$$

In the last equality we used (8.6.7) and the stationarity.

Consider the expectation on the right side of (8.6.8). Using stationarity of $\{X_t\}$ we get

$$\mathbb{E}\{\varphi_j(X_l)\varphi_j(X_t)\} = \mathbb{E}\{\varphi_j(X_0)\varphi_j(X_{|t-l|})\}. \tag{8.6.9}$$

Using this result and (8.2.14) in (8.6.8) yields,

$$\mathbb{V}(\hat{\theta}_j) = n^{-2}\sum_{l=-n}^{n}(n - |l|)\mathbb{E}\{\varphi_j(X_0)\varphi_j(X_l)\} - \theta_j^2. \tag{8.6.10}$$

The expression (8.6.10) allows us to make the following important conclusion. If the following sum converges,

$$\sum_{l=-n}^{n}|\mathbb{E}\{\varphi_j(X_0)\varphi_j(X_l)\}| < c^* < \infty, \tag{8.6.11}$$

then the variance of $\hat{\theta}_j$ decreases with the parametric rate n^{-1} despite the dependence between the observations, namely

$$\mathbb{V}(\hat{\theta}_j) < c^*n^{-1}. \tag{8.6.12}$$

Otherwise the dependence may slow down the rate of the variance decrease and hence make the estimation less accurate.

As an example, consider the case of a stationary series $\{X_t\}$ with the mixing coefficient $\alpha^X(s)$ defined in (8.1.10). Then using (8.1.13) we get

$$|\mathbb{E}\{\varphi_j(X_0)\varphi_j(X_s)\} - \theta_j^2| \le 8\alpha^X(s). \tag{8.6.13}$$

Using this relation we conclude that if the mixing coefficients are summable, that is

$$\sum_{s=0}^{\infty} \alpha^X(s) < \infty, \tag{8.6.14}$$

then (8.6.11) holds and we get the classical rate n^{-1} for the variance. If (8.6.14) holds then we are dealing with the case of *weak* dependency (*short memory*). For instance, if $\alpha^X(s)$ is proportional to $s^{-\beta}$ with $\beta > 1$ then this is the case of weak dependency. Another classical case of weak dependency is a Gaussian ARMA time series where mixing coefficients decrease exponentially. On the other hand, if $\beta < 1$, and an example will be presented shortly, the variance convergence slows down to $n^{-\beta}$. In the latter case the dependence is called strong and we may say that we are dealing with variables having *long memory* of order β.

Now we are explaining our approach for the case when the support of X is unknown and may be a real line like in the case of a Gaussian stationary time series. Denote by $X_{(1)}$ and $X_{(n)}$ the smallest and largest observations (recall that this is our traditional notation for ordered observations), define rescaled on $[0, 1]$ observations $Y_l := (X_l - X_{(1)})/(X_{(n)} - (X_{(1)})$, construct for the rescaled observations the density E-estimator $\hat{f}^Y(y)$, and then define the the rescaled back density E-estimator

$$\hat{f}^X(x) := \frac{\hat{f}^Y((x - X_{(1)})/(X_{(n)} - X_{(1)}))}{X_{(n)} - X_{(1)}} I(x \in [X_{(1)}, X_{(n)}]). \tag{8.6.15}$$

Let us explain the underlying idea of (8.6.15). Consider a variable Y' with density $f^{Y'}(y)$ supported on $[0, 1]$. Introduce a new variable $X' := a + bY'$ which is a scale-location transformation of Y'. Then the cumulative distribution function of X' is

$$\begin{aligned} F^{X'}(x) &= \mathbb{P}(X' \le x) = \mathbb{P}(Y' \le (x - a)/b) \\ &= F^{Y'}((x - a)/b) = \int_0^{(x-a)/b} f^{Y'}(y)dy. \end{aligned} \tag{8.6.16}$$

Note that the support of X' is the interval $[a, a + b]$, and hence via differentiation of (8.6.16) we get

$$f^{X'}(x) = b^{-1} f^{Y'}((x - a)/b) I(x \in [a, a + b]). \tag{8.6.17}$$

This is a classical formula for a scale-location transformation, and it motivated (8.6.15).

Now let us get a feeling of the effect of dependency on estimation of the density via simulated examples. Figure 8.7 allows us to understand performance of the E-estimator and complexity of the problem for the case of short-memory processes. The figure allows us to consider two ARMA(1,1) processes with different parameters. In the left column we consider the case of a highly oscillated zero-mean Gaussian ARMA(1,1) process $\{X_t\}$. In the right column we are also considering the case of a zero-mean Gaussian ARMA(1,1) process $\{Z_t\}$ only with parameters implying slower oscillations. Note that the horizontal dashed line in the top diagrams helps us to visualize oscillations around zero. Both $\{X_t\}$ and $\{Z_t\}$ are weak-dependent and short-memory processes with exponentially decreasing mixing coefficients, but they have different shapes of spectral densities, see Figures 8.1 and

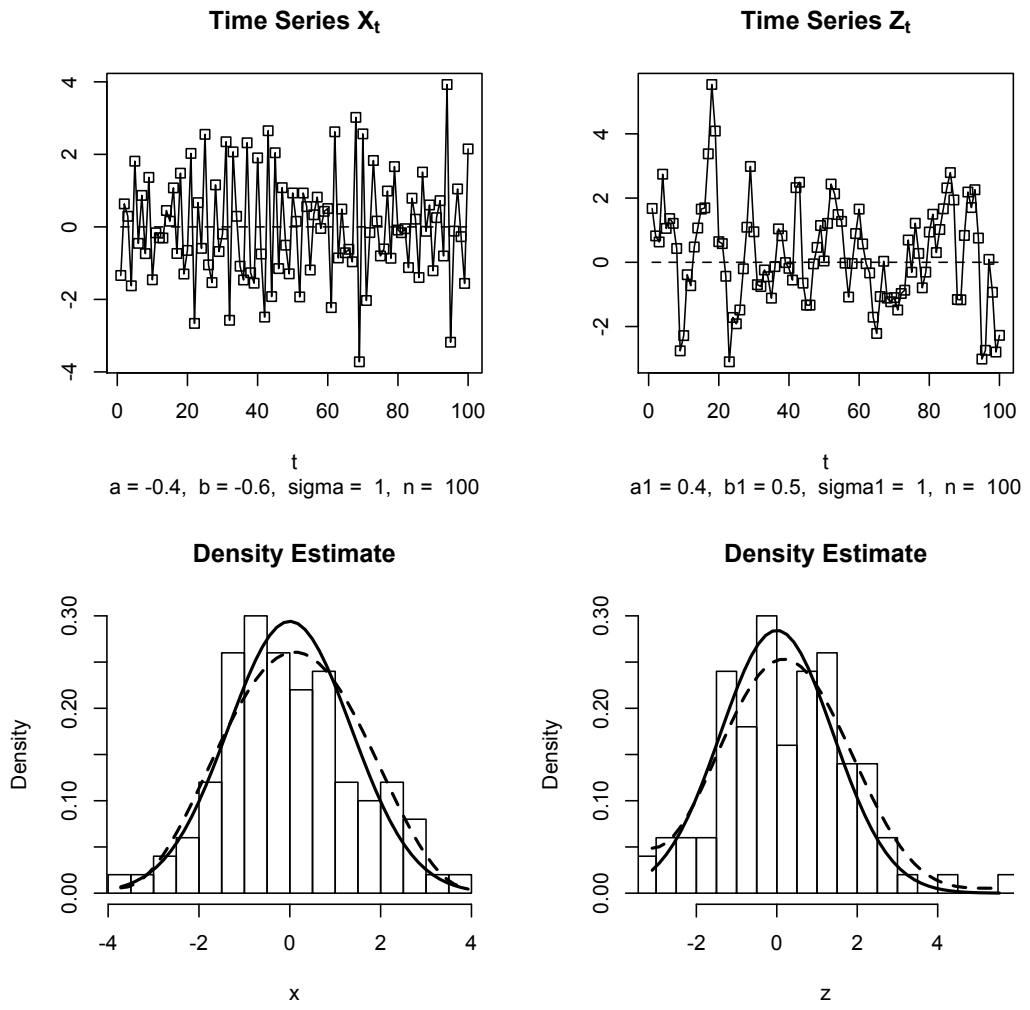

Figure 8.7 *Density estimation for short-memory stationary time series. Realizations of time series* $\{X_t\}$ *and* $\{Z_t\}$ *are generated by Gaussian ARMA(1,1) processes whose parameters are shown in the subtitles. Note that they are similar to those in Figures 8.1 and 8.2, respectively. The histograms are overlaid by the underlying density (the solid line) and the E-estimate (the dashed line) shown over the estimated support.* {*Parameters a, b and sigma control* $\{X_t\}$ *while a*1, *b*1 *and sigma*1 *control* $\{Z_t\}$.} *[n = 100, sigma = 1, a = -0.4, b = -0.6, sigma1 = 1, a1 = 0.4, b1 = 0.5 , cJ0 = 4, cJ1 = 0.5, cTH = 4]*

8.2, respectively. For each process, corresponding bottom diagrams shows us the histogram of available observations as well as the underlying Gaussian density and its E-estimate. Let us analyze and compare these diagrams. First of all, the left histogram is clearly more symmetric about zero than the right one. The right histogram is skewed to the right, and this reflects the fact that the observed series $\{Z_t\}$ spends more time above zero than below. This is because $\{Z_t\}$ has more power on lower frequencies and it requires more time (larger n) to exhibit its stationarity and zero-mean property. In terms of modes, the E-estimates indicate a pronounced single mode, but again the more frequently oscillating $\{X_t\}$ yields the better shape of the E-estimate. Finally, let us look at the effect of using the empirical support

(note that this is the same support used by the famous empirical cumulative distribution function $\hat{F}^X(x) := n^{-1}\sum_{l=1}^n I(X_l \le x)$). Of course, a Gaussian variable is supported on a real line, and here we used a finite empirical support. This did a good job for $\{X_t\}$ and a reasonable one for $\{Z_t\}$.

Now let us consider an even more extreme case of dependence, a long-memory time series. The top diagram in Figure 8.8 exhibits a zero-mean Gaussian time series with long memory of order $\beta = 0.4$. It looks like the above-discussed time series $\{Z_t\}$ on steroids. Note that it begins above the zero and rarely goes into negative territory (the horizontal dashed line helps us to see this). Of course, eventually the series will stay negative a long time, but much larger samples are needed to see this. The reader is advised to generate more series and get used to processes with long-memory. In the second (from the top) diagram we see the histogram of available observations overlaid by the underlying density (the solid line) and its E-estimate (the dashed line) shown over the range of observations. The data is clearly skewed to the right, the estimated support is skewed to the right, and there is nothing that can be done about this. We simply need more observations to correctly estimate the density, and this is an important lesson to learn.

What will be if some observations are missed? How does missing, coupled with dependency, affect E-estimation of the density? First, let us answer these natural questions analytically. Consider a time series $\{A_tX_t\}$ where the availabilities A_t are independent Bernoulli random variables and $\mathbb{P}(A_t = 1|\{X_t\}) = w$, $w \in (0, 1]$. Additionally, we are assuming that $\mathbb{P}(X_t = 0) = 0$ and hence $\mathbb{P}(A_t = I(A_tX_t \ne 0)) = 1$. As a result, even if we do not directly observe a particular A_t, we do know it from the available time series $\{A_tX_t\}$.

We are dealing with MCAR (missing completely at random) and hence it is natural to try a complete-case approach which yields the following Fourier estimator (compare with (8.6.4)),

$$\tilde{\theta}_j := \frac{\sum_{l=1}^n I(A_lX_l \ne 0)\varphi_j(A_lX_l)}{\sum_{l=1}^n I(A_lX_l \ne 0)}. \tag{8.6.18}$$

Let us check that the Fourier estimator is unbiased given $N := \sum_{l=1}^n A_l > 0$. Using the rule of calculation of an expectation via the expectation of a conditional expectation, together with $\mathbb{E}\{\varphi_j(X_l)\} = \theta_j$, we get

$$\begin{aligned}\mathbb{E}\{\tilde{\theta}_j|N > 0\} &= \mathbb{E}\Big\{\mathbb{E}\Big\{\frac{\sum_{l=1}^n I(A_lX_l \ne 0)\varphi_j(A_lX_l)}{\sum_{l=1}^n I(A_lX_l \ne 0)}|\{A_t\}, N > 0\Big\}|N > 0\Big\}\\ &= \mathbb{E}\Big\{\frac{\sum_{l=1}^n A_l\theta_j}{\sum_{l=1}^n A_l}|N > 0\Big\} = \theta_j.\end{aligned} \tag{8.6.19}$$

This is a pivotal result which yields unbiasedness of the sample mean estimator for the case of missing dependent observations. Hence we again may use our density E-estimator based on complete cases of time series $\{A_tX_t\}$.

Let us look at a simulation of a long-memory time series with missing observations. The third (from the top) diagram in Figure 8.8 shows us such a realization, and note that the time series of interest is shown in the top diagram. Clearly the density of available observations will be skewed to the right because there are only several negative observations. Further, note that we have dramatically less, just $N = 58$, available realizations of the long-memory time series. How will the E-estimator perform under these circumstances? The bottom diagram shows us the underlying density and the E-estimate. Yes, the estimate is skewed, and support is shown incorrectly, but overall this estimate is on par with its benchmark in the second (from the top) diagram. Despite all the complications, we do see a unimodal and surprisingly symmetric shape of the E-estimate. Note that the missing makes available observations less dependent, and then the main complication is the smaller number N of available observations.

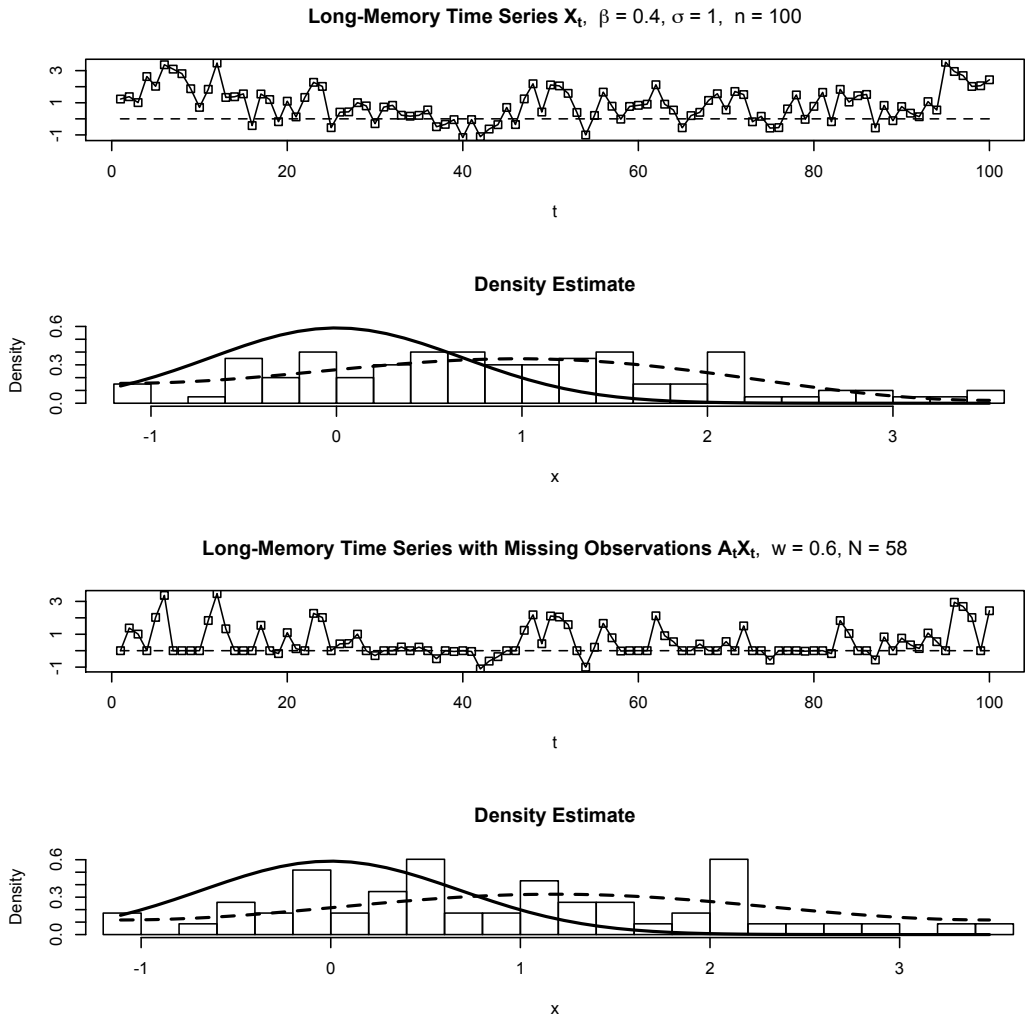

Figure 8.8 *Density estimation for long-memory of order β zero-mean Gaussian time series $\{X_t\}$ without and with missing observations. The missing is created by a time series $\{A_t\}$ of independent Bernoulli variables with $\mathbb{P}(A = 1) = w$, and then the available time series is $\{A_tX_t\}$. The title of the third diagram shows w and the number $N := \sum_{l=1}^{n} A_l$ of available observations. In the second and fourth diagrams the histograms are based on data in the first and third diagrams, respectively. The histograms are overlaid by solid and dashed lines showing the underlying density and its E-estimate over the range of available observations. {Parameter β is controlled by argument beta. Standard deviation of the time series is controlled by argument sigma.} [n = 100, sigma = 1, beta = 0.4, w = 0.6, cJ0 = 4, cJ1 = 0.5, cTH = 4]*

A conclusion is that dependency in observations should not be taken lightly, and typically larger sample size is the remedy. Further, while short-memory dependence rarely produces a dramatic effect on statistical estimation or inference, a long-memory dependence, coupled with a relatively small sample size, may be destructive. Working with and repeating Figures 8.7 and 8.8 will help the reader to understand and appreciate the dependency.

8.7 Nonparametric Autoregression

Consider a stationary time series $\{X_t\}$ generated according to the model

$$X_t = m(X_{t-1}) + \sigma(X_{t-1})W_t, \tag{8.7.1}$$

where $m(x)$ and $\sigma(x) > 0$ are smooth functions, and $\{W_t\}$ is a standard white noise, that is a time series of independent and identically distributed variables with zero mean and unit variance.

The time series (8.7.1) has many applications, interpretations, and it is known under different names. If in (8.7.1) $m(X_t) = aX_t$ and $\sigma(X_t) = \sigma$ then, according to (8.1.7), the process becomes AR(1) autoregression. This explains the name *nonparametric autoregression* for (8.7.1). Further, consider the problem of prediction (forecasting) X_t given X_{t-1}. For instance, we would like to predict the temperature for tomorrow based on the temperature today, or predict a stock return for tomorrow, etc. Then the best predictor, that minimizes the mean squared error, is $m(x) := \mathbb{E}\{X_t|X_{t-1} = x\}$ and it may be referred to as a nonlinear predictor or nonparametric regression. Further, in the theory of dynamic models, the equation is called a nonlinear dynamic model, X_t is called a state of the model, $m(x)$ is called an iterative map, and $s(x)$ is called a scale map. Note that if $\sigma(x) = 0$, then $X_t = m(X_{t-1})$ and a current state of this dynamic model is defined solely by its previous state (the states are iterated). This explains the name iterative map of $m(x)$.

Let us explain how we may estimate the autoregression function $m(x)$ in model (8.7.1). Set $Z_t := X_{t-1}$ and rewrite (8.7.1) as

$$X_t := m(Z_t) + \sigma(Z_t)W_t, \;\; t = 1, 2, \ldots, n. \tag{8.7.2}$$

Let us look at (8.7.2) more closely. First, we observe $n-1$ pairs $(X_2, Z_2), (X_3, Z_3), \ldots, (X_n, Z_n)$. Second, we can write using independence of $Z_t := X_{t-1}$ and W_t, together with the zero mean property of W_t,

$$\mathbb{E}\{X_t|Z_t = z\} = m(z) + \sigma(z)\mathbb{E}\{W_t\} = m(z). \tag{8.7.3}$$

We may conclude that (8.7.2) is a regression problem with Z_t being the predictor and X_t being the response. Hence we can use the regression E-estimator of Section 2.3 to estimate $m(z)$ and the scale E-estimator of Section 3.6 to estimate $\sigma(z)$.

Further, we can generalize model (8.7.1) and consider

$$X_t = m(X_{t-1}) + \sigma(X_{t-1})U_t, \tag{8.7.4}$$

where $\{U_t\}$ is a stationary and unit-variance time series satisfying $\mathbb{E}\{U_t|X_{t-1}\} = 0$. Then $\mathbb{E}\{X_t|X_{t-1} = x\} = m(x)$ and we again may use the regression and scale E-estimators.

Figure 8.9 presents the proposed statistical analysis of the nonparametric autoregression (8.7.4) with $\{U_t\}$ being a Gaussian ARMA(1,1) process; this process allows us to test robustness of the E-estimator to the assumption $\mathbb{E}\{U_t|X_{t-1}\} = 0$. The underlying simulation is explained in the caption. Diagram 1 shows us a particular realization. It is difficult to gain anything feasible from its visualization, and it looks like there is nothing special in this highly oscillated time series. Keeping in mind that the autoregression is often used to model a stock price over a short period of time, it is easy to understand why trading stocks is a complicated issue. Diagram 2 shows us the scattergram of pairs (X_{t-1}, X_t), it sheds light on the underlying iterative process of autoregression, and look at how inhomogeneous the scattergram is. The solid line is the underlying autoregression function $m(x)$, and it definitely sheds light on the data (of course, for real data this line would not be available). The interesting and rather typical feature is a small number of observations in the tails. Another interesting feature of the data is that the variability of observations depends on the predictor X_{t-1} and it is clearly larger for positive predictors.

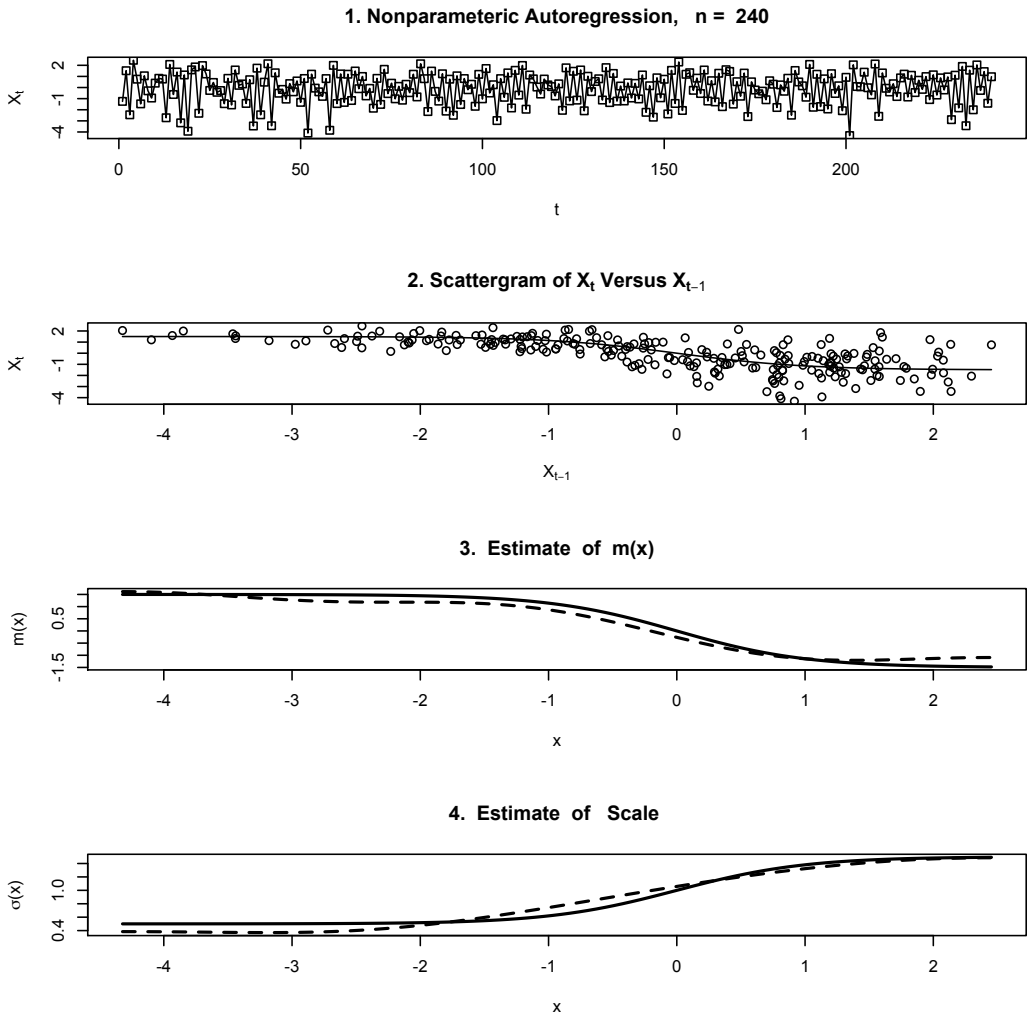

Figure 8.9 *Nonparametric autoregression. The simulation uses model (8.7.4) where* $m(x) := C\exp(\lambda_m x)/(1+\exp(\lambda_m x))$, $\sigma(x) := a_1 + b_1\exp(\lambda_s x)/(1+\exp(\lambda_s x))$, *and* $\{U_t\}$ *is a Gaussian ARMA(1, 1) time series with parameters* (a, b, σ). *In second and third diagrams the solid line is the underlying function* $m(x)$, *and in Diagram 4 the solid line is the underlying scale* $\sigma(x)$. *The dashed lines show E-estimates.* {*Parameters* λ_m *and* λ_s *are controlled by arguments lambdam and lambdas, respectively.*} *[n = 240, a = -0.3, b = 0.6, sigma = 1, a1 = 0.5, b1 = 1, lambdam = -2, lambdas = 2, C = 3, cJ0 = 4, cJ1= 0.5, cTH = 4]*

Now an important remark is due. In a nonparametric autoregression the role of noise $\{U_t\}$ is absolutely crucial because it forces $\{X_t\}$ to have a sufficiently large range of values which, in its turn, allows us to estimate the autoregression function. To appreciate the role of the noise, just set it to zero and then check the outcome theoretically and using Figure 8.9 (in the figure set $a1 = 0$ and $b1 = 0$).

Diagram 3 shows us the estimated autoregression function, and the E-estimate is good. Diagram 4 shows us the E-estimate of the scale function, and it is also good. Note that its left tail is smaller than the underlying scale but you can check diagram 2 and conclude that the data do support opinion of the E-estimate.

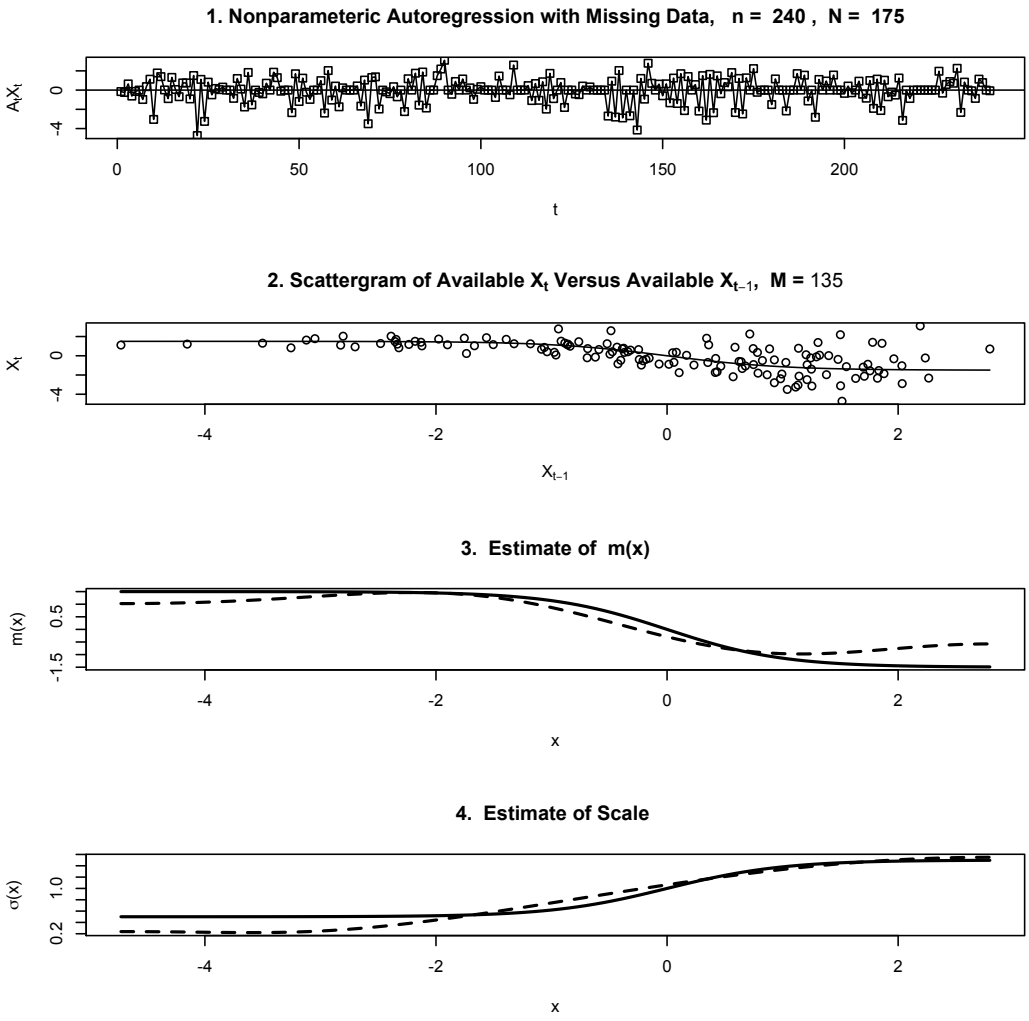

Figure 8.10 *Nonparametric autoregression with Markov-Bernoulli missing. Markov chain* $\{A_t\}$ *is generated according to transition probabilities* $\mathbb{P}(A_{t+1} = 0|A_t = 0) = \alpha$ *and* $\mathbb{P}(A_{t+1} = 1|A_t = 1) = \beta$*. Diagram 1 shows the observed time series* $\{A_tX_t\}$ *where* $\{X_t\}$ *is generated as in Figure 8.9.* $N := \sum_{l=1}^{n} A_l$ *is the number of available observations of* X_t *while* $M := \sum_{l=2}^{n} A_{l-1}A_l$ *is the number of available pairs of observations; these statistics are shown in the titles. [n = 240, alpha = 0.4, beta = 0.8, a = -0.3, b = 0.6, sigma = 1, a1 = 0.5, b1 = 1, lambdam = -2, lambdas = 2, C = 3, cJ0 = 4, cJ1 = 0.5, cTH = 4]*

Now let us consider the same model only when some observations are missing according to a Markov-Bernoulli time series $\{A_t\}$ discussed in Section 8.3. This model is illustrated in Figure 8.10 whose caption explains the simulation and notation. The underlying process $\{X_t\}$ is the same as in Figure 8.9. Diagram 1 shows us realization of the available time series $\{A_tX_t\}$ where only $N = \sum_{l=1}^{n} A_l = 175$ from $n = 240$ observations are available. Now note that E-estimation is based on pairs (X_{l-1}, X_l), $l = 2, \ldots, n$ and these pairs are available only if $A_{l-1}A_l = 1$. The available pairs are shown in diagram 2, and the number M of available pairs is 135. Note that we lost almost a half of underlying pairs (compare with Diagram 2 in Figure 8.9). This loss definitely has affected estimation of the autoregression

function (Diagram 3) and the scale function (Diagram 4). At the same time, with the help of Diagram 2 we may conclude that the estimates reflect the data. Indeed, let us look at the left tail in Diagram 2. Clearly all observations are below the solid line (the underlying autoregression), and this is what the E-estimate in diagram 3 tells us. Further, note that observations of X_t in the left tail of Diagram 2 exhibit a minuscular variability, and this is correctly reflected by the E-estimate in Diagram 4. It is possible to make a similar conclusion about the right tail.

We conclude that the proposed methodology of E-estimation is robust and performs relatively well for the complicated model of nonparametric autoregression with missing observations. The reader is advised to repeat Figures 8.9 and 8.10 with different parameters and get used to this important stochastic model.

8.8 Exercises

8.1.1 Consider a not necessarily stationary time series $\{X_t\}$ with a uniformly bounded second moment, that is $\mathbb{E}\{X_t^2\} \le c < \infty$ for any t. Is the first moment of X_t uniformly bounded?
8.1.2 Consider a stationary time series with a finite second moment. Does the first moment exist? If the answer is "yes," then can the first moment $\mathbb{E}\{X_t\}$ change in time?
8.1.3 Consider a zero-mean and second-order stationary time series $\{X_t\}$. Prove that the autocovariance function $\gamma^X(0) = \mathbb{E}\{X_t^2\} = \mathbb{V}(X_t)$ for any t.
8.1.4 Verify (8.1.3) and (8.1.4).
8.1.5* Show that for a short-memory second-order stationary time series the sample mean estimator of its mean has variance which vanishes with the rate n^{-1}.
8.1.6* Consider the case of a long-memory second-order stationary time series with $\gamma^X(t)$ proportional to $|t|^{-\alpha}$, $0 < \alpha < 1$. Evaluate the variance of the sample mean estimator $\bar{\mu} := n^{-1}\sum_{l=1}^n X_l$.
8.1.7 Let $\{W_t\}$ be a standard Gaussian white noise. For the following time series, verify the second-order stationarity and calculate the mean and autocovariance function:
(i) $X_t = a + bW_t$.
(ii) $X_t = a + bW_t \sin(ct)$.
(iii) $X_t = W_t W_{t-2}$.
(iv) $X_t = aW_t \cos(ct) + bW_{t-2}\sin(ct)$.
(v) $Z_t = a + bW_t + cW_t^2$.
8.1.8 Suppose that $\{X_t\}$ and $\{Z_t\}$ are two uncorrelated second-order stationary time series. What can be said about a time series $\{Y_t\} := a\{X_t\} + b\{Z_t\}$ where the sum is understood as the elementwise sum. Hint: Calculate the mean and autocovariance.
8.1.9 Prove that an autocovariance function always satisfies the inequality $\gamma^X(k) \le \gamma^X(0)$. Hint: Use Cauchy-Schwarz inequality.
8.1.10* Consider a realization $X_1, \dots, X_n$, $n \ge p$ of a causal AR(p) process (8.1.7). Prove that the estimator $\hat{X}_{n+1} := \sum_{k=1}^p a_k X_{n+1-k}$ is the best linear predictor of X_{n+1} that minimizes the mean squared error $\mathbb{E}\{(\tilde{X}_{n+1} - X_{n+1})^2\}$ over all linear estimators $\tilde{X}_{n+1} = \sum_{k=1}^n b_k X_{n+1-k}$. Hint: Write down the mean squared error and then minimize it with respect to $b_1, \dots, b_n$. Also, note that the mean squared error is not smaller than $\mathbb{E}\{W_{n+1}^2\}$.
8.1.11 What is the definition of a causal ARMA process? Why are these processes of interest?
8.1.12* Consider the process $X_t = X_{t-1} + W_t$. Show that it does not have a stationary solution.
8.1.13 Explain how MA(q) and AR(p) processes can be simulated.
8.1.14* Consider a Gaussian ARMA(1,1) process $\{X_t\}$ defined by $X_t - aX_{t-1} = \sigma(W_t + bW_{t-1})$ with $|a| < 1$ and $-a \ne b$. Show that the process is stationary and causal. Hint: Show that the process may be written as $X_t = \sigma W_t + \sigma(a+b)\sum_{j=1}^\infty a^{j-1} W_{t-j}$.

8.1.15 Consider the product of two independent second-order stationary time series one of which belongs to class (8.1.8). Does the product belong to class (8.1.8)?
8.1.16 For a second-order stationary times series $\{X_t\}$ from a class (8.1.8), what can be said about variance of a linear sum $\sum_{l=1}^{n} a_l X_l$? Hint: Introduce reasonable restrictions on numbers $a_l, l = 1, 2, \ldots, n$.
8.1.17 Consider a MA(2) process. What can be concluded about its mixing coefficient (8.1.10)?
8.1.18 Give an ARMA example of m-dependent time series.
8.1.19 Consider a stationary time series $\{X_t\}$ with a known mixing coefficient $\alpha^X(s)$. Evaluate the mean and variance of $Y := n^{-1} \sum_{l=1}^{n} \sin(X_l)$. Hint: Use (8.1.13) and any additional assumption needed.
8.1.20 Consider a stationary Gaussian ARMA process. What can be said about its mixing coefficient (8.1.10)? Hint: Use the Kolmogorov-Rosanov result.
8.1.21* Prove (8.1.12).
8.1.22* Verify (8.1.13).

8.2.1 Is the spectral density symmetric about zero (even function)? Also, calculate $\int_{-\pi}^{\pi} g^X(\lambda) d\lambda$.
8.2.2* Show that a spectral density is a nonnegative function.
8.2.3 How can Fourier coefficients of a spectral density $g^X(\lambda)$ be expressed via the autocovariance function?
8.2.4* Explain formula (8.2.3). Present a motivating example.
8.2.5 What is the difference, if any, between estimators (8.2.4) and (8.2.5)?
8.2.6 Compare biases of autocovariance estimators (8.2.4) and (8.2.5) for a zero-mean and second-order stationary time series.
8.2.7* Calculate variance and the mean squared error of the autocovariance estimator (8.2.4). Hint: Make your own assumptions.
8.2.8 Calculate variance of the estimator (8.2.5). Hint: Make your own assumptions.
8.2.9* Calculate the mean and variance of the periodogram.
8.2.10 Using your knowledge of nonparametric estimation, explain why a periodogram cannot be a consistent estimator.
8.2.11 What is the definition of a Gaussian ARMA(1,1) process?
8.2.12* Verify (8.2.7).
8.2.13 Prove that the autocovariance of an ARMA(1,1) process decreases exponentially.
8.2.14 In Figure 8.1 the ARMA process exhibits high fluctuations. Suggest parameters of an ARMA process that fluctuates slower. Use Figure 8.1 to verify your recommendation.
8.2.15 The periodogram in Figure 8.1 indicates a possible periodic (so-called seasonal) component in the underlying process. Use formula (8.2.3) to find the period of a possible periodic component. Do you believe that this component is present in the process shown in the top diagram?
8.2.16 Explain how the spectral density E-estimator is constructed.
8.2.17 Using Figure 8.1, propose better parameters of the E-estimator.
8.2.18 Using Figure 8.1, conduct simulations with different values of parameter σ, and report your findings.
8.2.19 Is it reasonable to believe that, as (8.2.8) indicates, the spectral density of an ARMA process is proportional to σ^2?
8.2.20 Explain relation (8.2.9).
8.2.21 In the right side of (8.2.11), one part is called the integrated variance (or simply variance), and another the integrated squared bias. Write down these two components and explain their names.
8.2.22 Prove (8.2.12).
8.2.23* Evaluate the variance of the sample autocovariance.

8.2.24 Verify (8.2.14). Hint: Write down the squared sum via a corresponding double sum, and then think about addends as elements of a matrix.
8.2.25 Verify (8.2.15).
8.2.26* Prove (8.2.16). Hint: Use the assumption that the time series is Gaussian.
8.2.27 Verify (8.2.17).
8.2.28 Why do we need the assumption (8.2.18)? Does it hold for ARMA processes?
8.2.29 Using Cauchy-Schwarz inequality, establish (8.2.19).
8.2.30 Prove (8.2.20).
8.2.31* Prove (8.2.21).
8.2.32 What do (8.2.21) and (8.2.22) tell us about the sample autocovariance estimator?
8.2.33* Use formula (8.2.23) and find optimal cutoff for an ARMA process. Then compare estimation of the spectral density with estimation of a single parameter.
8.2.34 Verify (8.2.25).
8.2.35 Explain how (8.2.28) is obtained and why the proposed estimator $\bar{g}^X$ presents a practical interest.
8.2.36* Suppose that (8.2.26) is correct. The aim is to propose an estimator and to choose a minimal sample size that the MISE does not exceed a fixed positive constant ε. Propose such an estimator and the sample size.

8.3.1 Explain the missing mechanism (8.3.1).
8.3.2 In the case of a time series with missing observations, we observe realizations of two time series $\{A_tX_t\}$ and $\{A_t\}$. Explain why under the assumption $\mathbb{P}(X_t = 0) = 0$ it is sufficient to know realizations of only one time series $\{A_tX_t\}$.
8.3.3 Prove that if $\mathbb{P}(X_t = 0) = 0$, then $\mathbb{P}(A_t = I(Y_t \neq 0)) = 1$ where Y_t is defyned in (8.3.1). Does this imply that $A_t = I(Y_t \neq 0)$?
8.3.4 Suppose that the chance of rain tomorrow depends only on rain or no rain today. Suppose that if it rains today, then tomorrow it will rain with probability α, and if it is no rain today, then tomorrow it will rain with probability β. Find the probability that if today is no rain, then two days from today there will be rain. Hint: Consider a two-state Markov chain.
8.3.5* For the previous problem, consider 10 consecutive days and find the expected number of rainy days.
8.3.6* Consider a Markov-Bernoulli missing mechanism with $\alpha := \mathbb{P}(A_{t+1} = 0|A_t = 0)$. Let L be the length of a batch of missing cases. Explain why, given the batch length $L \geq 1$, the distribution of L is geometric with $\mathbb{P}(L = k) = \alpha^{k-1}(1 - \alpha)$, $k = 1, 2, 3, \ldots$
8.3.7* For the setting of Exercise 8.3.6, find the mean and the variance of the batch length L.
8.3.8 Is the autocovariance $\gamma^A(j)$ equal to $\mathbb{E}\{A_1A_{1+j}\}$?
8.3.9 Give the definition of a Markov chain.
8.3.10 Is an ARMA(1,1) process a Markov chain? If the answer is "yes," then what is the order of the Markov chain? Hint: Think about the effect of parameters of an ARMA processes.
8.3.11 Explain the simulation used to create Figure 8.2.
8.3.12 Repeat Figure 8.2 for different Markov-Bernoulli processes. Explain how its parameters affect the missing and estimation of the spectral density.
8.3.13* Why does the naïve estimate in Figure 8.2 indicate a lower (with respect to the E-estimate) spectrum power on low frequencies?
8.3.14 Explain the three estimates shown in Figure 8.2.
8.3.15 Using repeated simulations of Figure 8.2, propose better parameters of the E-estimator.
8.3.16* Find the mean and variance of the available number $N = \sum_{l=1}^{n} A_l$ of observations.
8.3.17 Verify (8.3.3).

8.3.18 Explain every equality in (8.3.4). Do not forget to comment on used assumptions.
8.3.19 What is the underlying idea of the estimator (8.3.5)? Explain its numerator and denominator.
8.3.20 Show that the autocovariance estimator (8.3.5) is unbiased.
8.3.21* Calculate the variance of estimator (8.3.5). Hint: Propose your assumptions.
8.3.22 Explain the underlying simulation in Figure 8.3.
8.3.23 Explain how parameter λ affects the number N of available observations in the simulation of Figure 8.3.
8.3.24 Formulate basic statistical properties of a batch-Bernoulli process.
8.3.25* Explain how the three estimates, shown in Figure 8.3, are constructed.
8.3.26* Given the same number of available observations, is Markov-Bernoulli or batch-Bernoulli missing mechanism better for estimation? You may use either a theoretical approach or simulations to answer the question.
8.3.27 How many parameters are needed to define a stationary Markov chain of order 2? Hint: It may be helpful to begin with a Markov chain of order 1.

8.4.1 Explain an amplitude-modulated missing mechanism. Give several examples.
8.4.2 Suppose that $\{X_t\}$, $\{A_t\}$ and $\{U_t\}$ are (second-order) stationary time series. Is their product a (second-order) stationary time series?
8.4.3 What do we need the assumption (8.4.1) for?
8.4.4* Consider the case when $\{A_t\}$ and $\{U_t\}$ are dependent. Does this affect estimation of the spectral density g^X?
8.4.5* Consider the case when $\{A_t\}$ and $\{X_t\}$ are dependent. Does this affect estimation of the spectral density g^X?
8.4.6* Find the mean and variance of the sample mean autocovariance (8.4.2).
8.4.7 Verify each equality in (8.4.3). Explain where and how the made assumptions about the three processes are used.
8.4.8 Can the spectral density g^X be consistently estimated based on amplitude-modulated observations? In other words, is this missing destructive?
8.4.9* Give definition of the shape of a function. Then explain when and how shape of the spectral density may be estimated for amplitude-modulated data.
8.4.10* Find the mean and variance of estimator (8.4.6).
8.4.11 Verify and explain all steps in establishing (8.4.7).
8.4.12 Explain the simulation used in Figure 8.4.
8.4.13 Explain how parameters of the processes $\{A_t\}$ and $\{U_t\}$ affect the number N of available observations. Support your conclusion using Figure 8.4.
8.4.14 Find better parameters of the E-estimator used in Figure 8.4.
8.4.15 Explain why the naïve estimate in Figure 8.4 indicates a smaller spectrum power on low frequencies.
8.4.16 Explain the model of amplitude-modulation by a Poisson variable.
8.4.17 In general, an amplitude-modulation implies inconsistent estimation of the spectral density. On the other hand, the Poisson amplitude modulation does allow a consistent estimation. Why?
8.4.18* Explain the estimator (8.4.8) of the mean of a Poisson distribution. Then evaluate its mean and variance.
8.4.19* Prove (8.4.9). Explain the used assumption.
8.4.20* Find the mean and variance of the estimator (8.4.10).
8.4.21* Consider the case when $\{X_t\}$ and Poisson $\{U_t\}$ are dependent. Explore the possibility of a consistent estimation of the spectral density or its shape.
8.4.22 Explain the simulation used in Figure 8.5.
8.4.23 Find the mean and variance of the available number N of observations in Figure 8.5. Does an underlying (hidden) time series $\{X_t\}$ affect N?

8.4.24 Consider the bottom diagram in Figure 8.5. The naïve estimate exhibits a larger spectrum power at all frequencies. Why?
8.4.25 Use Figure 8.5 to answer the following question. How do parameters of the underlying ARMA process affect estimation of its spectral density?

8.5.1 Explain the model of right censored time series. Present examples.
8.5.2 What are the available observations when time series is censored?
8.5.3 Explain formula (8.5.1).
8.5.4 What is the definition of a bivariate survival function? What are its properties?
8.5.5 Prove (8.5.3).
8.5.6* Find the mean and variance of estimator (8.5.4). Is it unbiased?
8.5.7* Explain the method of estimation of the survival function $G^C(v)$.
8.5.8 Verify (8.5.5).
8.5.9 Explain why the formula (8.5.6) is of interest.
8.5.10 What is the definition of a hazard rate? What are its properties?
8.5.11 Assume that the hazard rate is known. Suggest a formula for the corresponding probability density.
8.5.12 Verify (8.5.9).
8.5.13* Explain how the numerator and denominator in (8.5.9) may be estimated.
8.5.14 Prove validity of (8.5.11).
8.5.15* Find the mean and variance of the estimator (8.5.12).
8.5.16* Use an exponential inequality to infer about estimator (8.5.12).
8.5.17* Evaluate the mean and variance of estimator (8.5.13). Is it unbiased? Is it asymptotically unbiased?
8.5.18 Prove (8.5.14).
8.5.19 Explain why (8.5.15) is a reasonable estimator of the survival function.
8.5.20 Explain the simulation used to create Figure 8.6.
8.5.21 Explain all diagrams in Figure 8.6.
8.5.22 Consider Figure 8.6 and answer the following question. Why is the censored time series highly oscillated, while the underlying $\{X_t\}$ is not?
8.5.23 Explain how the naïve spectral density estimate is constructed.
8.5.24 The considered problem is complicated. Repeat Figure 8.6 a number of times and make your own conclusion about the E-estimator.
8.5.25 Use Figure 8.6 and then explain your observations about the size N of uncensored observations.
8.5.26 Suggest better parameters for the E-estimator. Hint: Use Figure 8.6 with different sample sizes and parameters.
8.5.27* Consider the case of a stationary time series $\{C_t\}$. Propose a consistent estimator of the bivariate survival function $G^{C_1,C_{1+j}}(v,u)$.
8.5.28* Consider the same setting only for time series of lifetimes. Propose a spectral density estimator and justify its choice. Hint: Estimate the mean, subtract it, and then check how this step affects statistical properties of the E-estimator.

8.6.1 Give definition of the probability density $f^X(x)$ of a continuous random variable X.
8.6.2 Suppose that $[0,1]$ is the support for a random variable X (or we may say the support of the probability density f^X). What is the meaning of this phrase?
8.6.3 Find the mean and variance of the sample mean estimator (8.6.4) for the case of independent observations of (a sample from) X.
8.6.4 Consider a time series $\{X_t\}$. What is the assumption that allows us to define the density f^{X_t}? What is the assumption that makes feasible the problem of estimation of the density f^{X_t}?
8.6.5 Explain why (8.6.6) is still valid for the case of dependent observations.

8.6.6 Is the sample mean estimator $\hat{\theta}_j$ unbiased? Is it robust toward dependence between observations?
8.6.7* Find the variance of the sample mean estimate $\hat{\theta}_j$ for the case of dependent observations. Explain all steps and made assumptions.
8.6.8 Explain all steps in establishing (8.6.8). Write down all used assumptions.
8.6.9 What is the assumption needed for validity of (8.6.9)?
8.6.10 Explain how the equality (8.6.10) was obtained.
8.6.11 Why is the assumption (8.6.11) important?
8.6.12 Explain importance of conclusion (8.6.12) for estimation of the probability density.
8.6.13 Give a definition of a weak dependence. Compare with the case of processes with long memory.
8.6.14 Suppose that (8.6.14) holds. What can be said about dependency between observations?
8.6.15 Consider a continuous random variable Z with density f^Z. What is the density of $Y := aZ + b$? Note that we are dealing with a scale-location transformation of Z.
8.6.16 Explain the motivation behind estimator (8.6.15). Why do we use such a complicated density estimator?
8.6.17* The density estimator (8.6.15) tells us that the support of X is $[X_{(1)}, X_{(n)}]$. Is this also the case for the empirical cumulative distribution function?
8.6.18* Consider a sample of size n from X. Find the probability $\mathbb{P}(X \geq X_{(n)})$. Use your result to improve the approach (8.6.15).
8.6.19 Explain formulas (8.6.16) and (8.6.17).
8.6.20 Explain the simulation that creates top diagrams in Figure 8.7.
8.6.21 Explain how estimates, shown in Figure 8.7, are calculated.
8.6.22 What is the difference, if any, between left and right columns of diagrams in Figure 8.7?
8.6.23 Repeat Figure 8.7 a number of times, and make a conclusion about which type of ARMA processes benefits estimation of the density.
8.6.24 Repeat Figure 8.7 with different parameters σ and σ_1. Report on how they affect estimation of the density.
8.6.25 Find better parameters of the E-estimator used in Figure 8.7.
8.6.26 Explain the simulation that creates the top diagram in Figure 8.8.
8.6.27 Explain the estimate shown in the second from the top diagram in Figure 8.8.
8.6.28 Explain the simulation that creates the third from the top diagram in Figure 8.8.
8.6.29 Explain the estimate shown in the bottom diagram in Figure 8.8.
8.6.30 Evaluate the mean and variance of the available number N of observations in simulation in Figure 8.8.
8.6.31 Suggest better parameters of the E-estimator used in Figure 8.8.
8.6.32* Using Figure 8.8, as well as your understanding of the theory, comment on the effect of missing data on density estimation for processes with long memory. Hint: Think about the case when the sample size of a hidden sample and the number of complete cases in a larger sample with missing observations are the same.
8.6.33 Explain the underlying idea of the estimator (8.6.18).
8.6.34* Find the mean and variance of the estimator (8.6.18).
8.6.35* Suppose that you have a sample of observations. What test would you suggest for independence of observations?
8.6.36* Propose a generator of a time series with long memory.

8.7.1 Explain the model of nonparametric autoregression.
8.7.2 Why can the model (8.7.1) be referred to as the prediction (forecasiting) model?
8.7.3* Consider a model $X_t = m_1(X_{t-1}) + m_2(X_{t-2}) + \sigma(X_{t-1}, X_{t-2})W_t$. Propose E-estimators for functions $m_1(x)$, $m_2(x)$ and $\sigma(x_1, x_2)$.

8.7.4 Is the process, defined by (8.7.1), second-order stationary? Hint: Think about assumptions.
8.7.5 Explain (8.7.2).
8.7.6 Verify (8.7.3).
8.7.7 Explain how the nonparametric autoregression model is converted into a nonparametric regression model.
8.7.8 Describe the simulation used in Figure 8.9.
8.7.9 How does a Markov-Bernoulli missing mechanism perform?
8.7.10 Is there any useful information that may be gained from analysis of Diagram 1 in Figure 8.9?
8.7.11 How was Diagram 2 in Figure 8.9 created?
8.7.12* Explain all steps in construction of E-estimator of $m(x)$. Then conduct a number of simulations, using Figure 8.9, and comment on performance of the estimator.
8.7.13* Use the theory and Figure 8.9 to explain how parameters of the ARMA process affect estimation of $m(x)$. Hint: Check the assumption $\mathbb{E}\{U_t|X_{t-1}\} = 0$.
8.7.14 Scale function in model (8.7.1) is an important function on its own, and it is often referred to as the volatility. Explain how it may be estimated using E-estimator. Then check its performance using Figure 8.9.
8.7.15 Repeat Figure 8.9 several times, make hard copies of figures, and then write down a report that explains performance of the E-estimators.
8.7.16 Find better parameters of the E-estimator used in Figure 8.9.
8.7.17 Explain how parameters of the E-estimator affect estimation of $m(x)$. Then test your conclusion using Figure 8.9.
8.7.18 Explain how parameters of the E-estimator affect estimation of $\sigma(x)$. Then test your conclusion using Figure 8.9.
8.7.19* Using Figure 8.9, find how parameters of the underlying model for $m(x)$ affect estimation of $m(x)$. Then explain your conclusion theoretically.
8.7.20* Using Figure 8.9, find how parameters of the underlying model for $m(x)$ affect estimation of $\sigma(x)$. Then explain your conclusion theoretically.
8.7.21 Using your understanding of the theory and simulations conducted by Figure 8.10, explain how parameters of the Markov-Bernoulli missing affect estimation of $m(x)$.
8.7.22 Scale function in model (8.7.1) is an important function on its own, and it is often referred to as the volatility. Explain how it may be estimated using E-estimator.
8.7.23 Using your understanding of the theory and simulations conducted by Figure 8.10, explain how parameters of the Markov-Bernoulli missing affect estimation of $\sigma(x)$.
8.7.24 Repeat Figure 8.10 several times, make hard copies of figures, and then write a report which explains shapes of the recorded E-estimates.
8.7.25* Model (8.7.1) is often referred to as a nonlinear one-step prediction. What will be the definition of a nonlinear two-step prediction model? Suggest an E-estimator.
8.7.26* Consider a functional-coefficient autoregression model

$$X_t = a_1(X_{t-d})X_{t-1} + \ldots + a_p(X_{t-d})X_{t-p} + \sigma(X_{t-d})W_t. \tag{8.8.1}$$

Propose an E-estimator for the functions $a_1(x), \ldots, a_p(x)$.

8.9 Notes

There is a number of excellent books devoted to time series analysis, for example Anderson (1971), Dzhaparidze (1985), Diggle (1990), Brockwell and Davis (1991), Fan and Yao (2003, 2015), Bloomfield (2004), Györfi et al. (2013), Box et al. (2016), Montgomery, Jennings and Kulahci (2016), Shumway and Stoffer (2017). The last three books contain the introduction to R and many R-examples. A large number of examples in applied economics can be found

in the book Greiner, Semmler and Gong (2005). A book-length treatment of hidden Markov models for time series using R can be found in Zucchini, MacDonald and Langrock (2016). Chapter 5 in Efromovich (1999a) is devoted to a number of topics in nonparametric time series analysis, and the review of classical time series results is based on that chapter.

8.1 Discussion of ARMA processes, function class (8.1.8) and a wide spectrum of related results can be found in Chapter 3 of Brockwell and Davis (1991), Efromovich (1998b, 1999a), Montgomery, Jennings and Kulahci (2016). Mixing processes and weak dependence are discussed in Doukhan (1994) and Dedecker et al. (2007).

8.2 The spectral density is defined via a cosine series approximation (8.2.2), and hence the use of E-estimator becomes natural and attractive. For all other applications, like estimation of regression, hazard rate or probability density, one may argue for and against a cosine-series approach in particular and an orthogonal series approach in general (versus other approaches like kernel or spline). And nonetheless, kernel smoothing of the periodogram was the first proposed and so far the most popular method of estimation of the spectral density. Lag-window spectral density estimates are discussed in Wu and Zaffaroni (2017) where further references may be found. Parts of Sections 8.1 and 8.2, devoted to a review of classical results of the time series theory, are based on Chapter 5 in Efromovich (1999a), Efromovich and Pinsker (1981) and Efromovich (1984), where more discussion and examples may be found. Asymptotic justification of the series estimation methodology can be found in Levit and Samarov (1978), Efromovich and Pinsker (1981, 1986), and Efromovich (1984, 1998b, 1999a).

8.3 The problem of nonparametric spectral density estimation for discrete time series in the presence of missing observations has a long history. Parzen (1963) has coined the term amplitude-modulated for the time series $\{A_tX_t\}$ and the term amplitude-modulating for the time-series $\{A_t\}$. Among first and most prominent theoretical results devoted to the spectral density estimation with stochastically missing observations, let us mention Scheinok (1965) who studied the case of A_t, $t = 1, 2, \dots$ being independent and identically distributed Bernoulli random variables, and Bloomfield (1970) who considered the case of a stationary $\{A_t\}$. For these settings, consistent spectral density estimators have been proposed. Further references and developments in time-series analysis with missing data can be found in Dunsmuir and Robinson (1981a,b), Baisch and Bokelmann (1999), Jiang and Hui (2004), Lee (2004), Tarczynski and Allay (2004), Robinson (2008), Vorotniskaya (2008), Baby and Stoica (2010), Butcher and Gillard (2016), and Jiang et al. (2016). The book by Ross (2014) gives a good introduction to Markov chains.

First consistent estimators of the spectral density have been developed at about the same time as consistent estimators for nonparametric regression. At the same time, while the theory of efficient and adaptive nonparametric regression and probability density estimation is known for more than four decades, the theory of efficient estimation of the spectral density with missing data was developed only recently, see Efromovich (2014b, 2017).

8.4 Examples of amplitude-modulated times series may be found in Parzen (1963), Dunsmuir and Robinson (1981a,b), Efromovich (1999a), Bloomfield (2004), Jiang and Hui (2004), and Vorotniskaya (2008). The asymptotic analysis of the proposed estimation procedure, which proves efficiency of the E-estimation approach, may be found in Efromovich (2014d). An interesting and practically important topic is the sequential estimation with assigned risk, see a discussion and corresponding results in Efromovich (2016b).

8.5 An interesting analysis of econometric methods and examples can be found in Li and Racine (2007) and Robinson (2008). Helsel (2011) provides a number of environmental examples and a discussion of Minitab and R softwares. Another good time series book to read is Box et al. (2016). Sun and Zhao (2013) provides a book-length treatment of survival topics including analysis of interval-censored failure time data. Among recent papers devoted to the analysis of censored time series, a special case of censored autoregressive model is considered in Wang and Chan (2017) where an elaborate system of unbiased estimating

equations is proposed, see also Choi and Portnoy (2016) where quantile autoregression is explored for censored time series data.

8.6 The book by Beran (1994) is devoted to long-memory processes. Probability density estimation and conditional density estimation for dependent observations is considered in the book by Fan and Yao (2003) where a local polynomial technique is used.

8.7 Nonparametric autoregression is a classical topic in nonlinear time series. Books by Fan and Gijbels (1996), Efromovich (1999a), Fan and Yao (2003, 2015) and Tsay (2005) discuss different aspects of this topic and present interesting examples and extensions. The discussion of nonparametric functional autoregression may be found in Zhu and Politis (2017).