Statistical Modeling with Spline Functions Methodology and Theory

Mark H. Hansen
University of California at Los Angeles

Jianhua Z. Huang
University of Pennsylvania

Charles Kooperberg
Fred Hutchinson Cancer Research Center

Charles J. Stone
University of California at Berkeley

Young K. Truong
University of North Carolina at Chapel Hill

January 5, 2006

8
Estimation of the Spectral Distribution

8.1 An example

8.1.1 The network data

In Figure 8.1 we show a time series of network data from the file server of the math center of Bell Labs covering the 18 week period from 5 am on August 27 to 5 am on December 31 1998. Each point represents how many kByte went into the file server during a five minute interval. Already this rough graph suggests that there may be a periodicity of a week in the data. The autocorrelation function in Figure 8.2 shows a clear periodicity of one day and a second periodicity of one week. In Figure 8.3 we look at the 126 daily total amounts of data that was going into the file server during this period. In Figure 8.4 we plot the average amount of data that was going into the file server during a particular five minute interval of any week. We again recognize the daily and weekly patterns in these figures. As a comparison, Figure 8.5 and 8.6 show the same summaries as Figure 8.3 and 8.4 respectively, for the data that was going out of the file-server. Here we also see a daily pattern, as well as a slight increase of the amount of data that was going out over time, but a weekly pattern is less clear. The difference between the graphs for the data going into and going out of the file server leads us to believe that a system of daily and maybe weekly backups is being used. Both the graphs for data going into and going out of the file server have a strong periodicity, but the volume of data that is going into the file server close to midnight of every day except Friday is much higher than volumes seen at any other time. The graphs of data

both going into and going out of the file-server also show some short time dependence: if the amount of data going into (out of) the file server is large during any five minute interval, it is likely also to be large during the next five minute interval.

Rather than analyzing the network data using fairly ad hoc techniques, as done above, we can also try to model the corresponding spectral distribution. In this chapter we will discuss the Lspec methodology for modeling a possibly mixed spectral distribution.

8.1.2 Background

The problem of estimating the spectral distribution for a stationary time series is of fundamental importance in statistics. If this distribution is absolutely continuous, its density function can be estimated by a variety of methods, the most popular being window, AR, and ARMA estimates. Recently wavelet methods have been proposed for this problem.

Window estimates are obtained by smoothing the periodogram using "window" (kernel) functions, while AR and ARMA estimates are obtained by fitting parametric AR and ARMA models using "automatic" model selection procedures such as AIC and BIC; see Priestley (1981, Chapters 6 and 7).

It is known that the periodogram is not a consistent estimate of the spectral density function and that consistency can be achieved by smoothing the periodogram ordinates, the degree of smoothing being controlled by the window width. Larger window widths smooth out the noise, but also tend to distort the details of the signal, while smaller window widths tend to yield estimates with spurious features.

In the AR approach the estimated spectral density function has the parametric form of the spectrum of an autoregressive process, and the resulting fit is better when the spectral density function can accurately be approximated by such a form. However, this procedure may yield poor estimates when it is used to fit simple MA models (Beamish and Priestley 1981).

The ARMA approach extends the AR approach by approximating the spectral density function with the spectrum of an ARMA model. Since many more spectral density functions can accurately be approximated by ARMA models, not surprisingly, ARMA estimates are often better than AR estimates in fitting non-AR models. However, in our experience, numerical procedures for determining maximum likelihood estimates of ARMA parameters are typically much less stable and far more computer intensive than the other approaches discussed in this chapter, making the ARMA approach considerably less attractive (see Section 7 of Kooperberg, Stone, and Truong 1995c).

Wavelet methods for estimation of the spectral density usually also start with the periodogram. Moulin (1994) and Gao (1997) use wavelet thresholding methods, similar to those proposed by Donoho and Johnstone (1994),

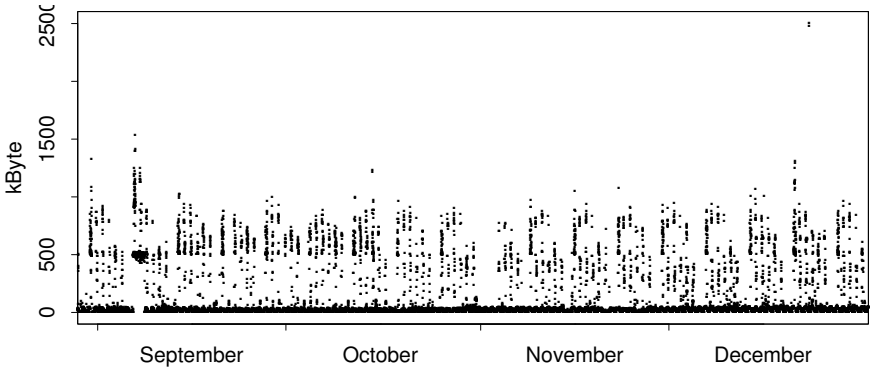

FIGURE 8.1. Amount of data that going into the file server of the math center of Bell Labs in five minutes intervals between August 27 and December 31, 1998.

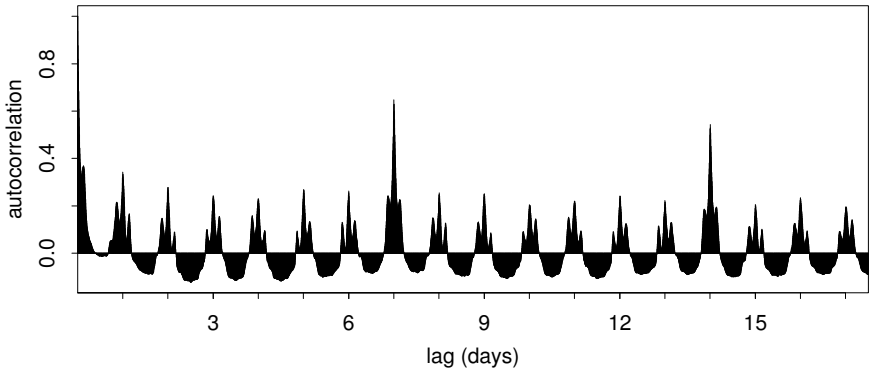

FIGURE 8.2. Autocorrelation function for the data shown in Figure 8.1.

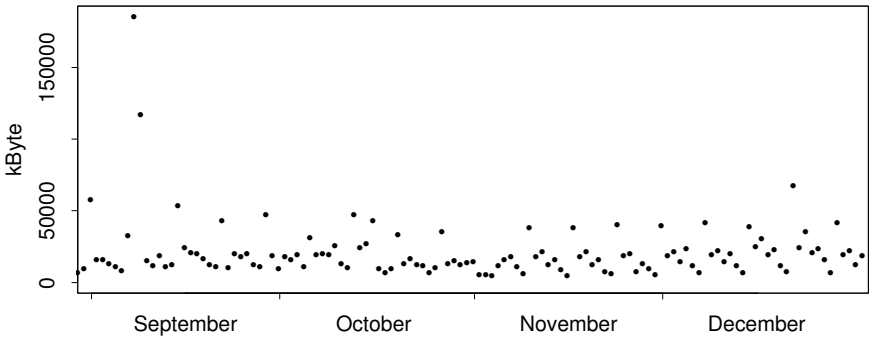

FIGURE 8.3. Amount of data that was going into the file server of the math center of Bell Labs in daily intervals between August 27 and December 31, 1998.

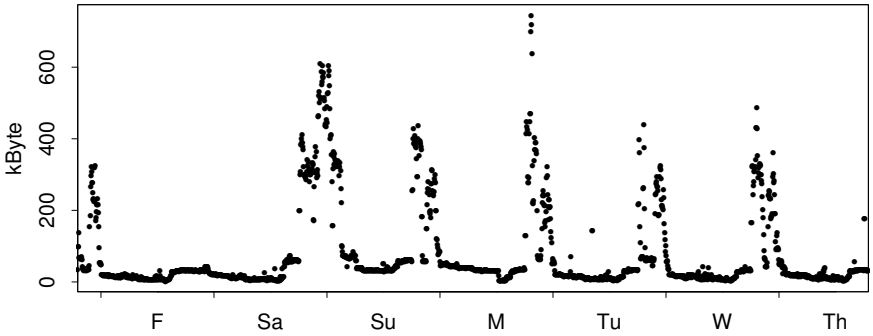

FIGURE 8.4. Average amount of data that was going into the file server of the math center of Bell Labs in five minute intervals during fixed times of the week over the last 18 weeks of 1998.

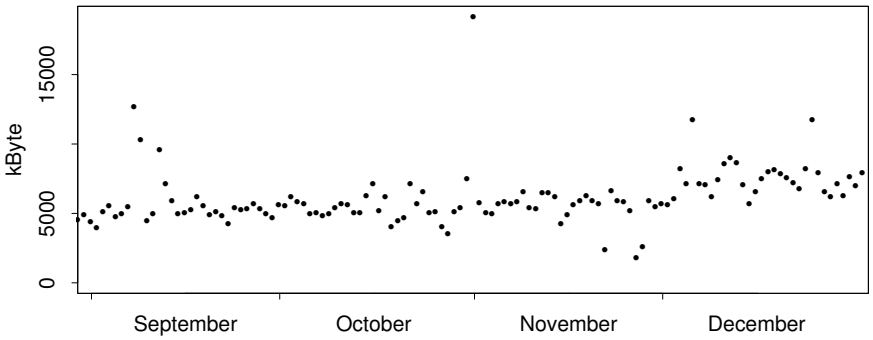

FIGURE 8.5. Amount of data that was going out of the file server of the math center of Bell Labs in daily intervals between August 27 and December 31, 1998.

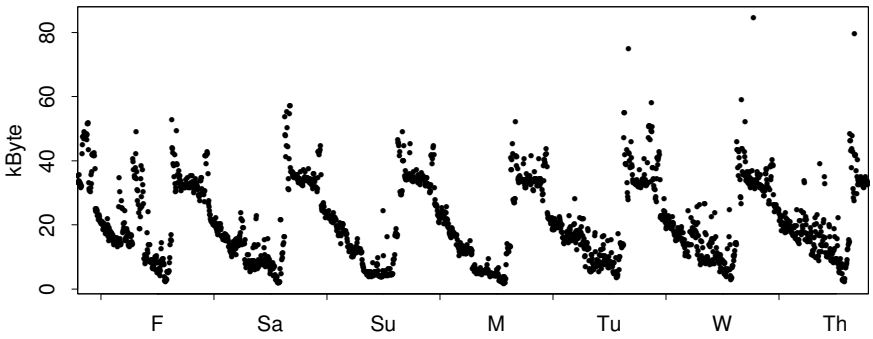

FIGURE 8.6. Average amount of data that was going out of the file server of the math center of Bell Labs in five minute intervals during fixed times of the week over the last 18 weeks of 1998.

that take the nonnormality of the periodogram into account. Rather than starting with the periodogram, Walden, Percival, and McCoy (1998) start with a multitaper estimate of the log-spectrum, which by itself is already more approximately normal.

If the spectral distribution is possibly mixed, there are two general approaches to its estimation that have previously been discussed. One approach is to apply a method that was actually designed for the absolutely continuous case and hope that the estimate has sharp peaks centered near the atoms. The AR methods are generally used for this purpose since they typically yield sharper peaks corresponding to atoms than window estimates, while they are numerically more stable than ARMA estimates. Mackisack and Poskitt (1990) give an asymptotic justification for the AR procedure. In the engineering literature there have been various proposals to make AR methods more sensitive to atoms in the spectral distribution (see, for example, Stoica, Moses, Söderström, and Li 1991).

The alternative approach is first to test whether atoms are present. If so, their locations and masses are estimated, the corresponding components are filtered out, and the spectral density function is estimated from the filtered time series. Two early references on this approach are Priestley (1962a) and Priestley (1962b). See Priestley (1981; Chapter 8) for an overview.

The Lspec methodology is an automatic procedure for estimating a possibly mixed spectral distribution. This method has the advantage of AR and ARMA estimates in being automatic and the advantage of the test-based procedures in giving explicit estimates for the masses and locations of the atoms. In this procedure the logarithm of the spectral density function is modeled as a polynomial spline, the unknown parameters of which are estimated by maximizing an approximation to the log-likelihood function.

8.1.3 Mixed Spectra

Consider a real-valued second order stationary time series X_t with mean $E(X_t) = E(X_0)$ and covariance function $\gamma(u) = \mathrm{cov}(X_t, X_{t+u})$. Assume that the time series has the form

$$X_t = \sum_{j=1}^{p} R_j \cos(t\lambda_j + \phi_j) + Y_t. \tag{8.1.1}$$

Here $0 < \lambda_j \leq \pi$; ϕ_j are independent and uniformly distributed on $[-\pi, \pi]$; R_j are independent, non-negative random variables such that R_j^2 has positive mean $4\rho_j$; and Y_t is a second-order stationary time series such that $E(Y_t) = E(X_0)$ and

$$\sum_u |\gamma_c(u)| < \infty, \tag{8.1.2}$$

where $\gamma_c(u) = \mathrm{cov}(Y_t, Y_{t+u})$.

For a time series X_t that satisfies these conditions, we now first define several functions related to the spectral distribution. In particular, the *spectral density function* of the time series is given by

$$f_c(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma_c(u) \exp(iu\lambda), \qquad -\pi \le \lambda \le \pi, \tag{8.1.3}$$

which can be extended to $(-\infty, \infty)$ in the obvious manner so as to be periodic with period 2π; its *line spectrum* is given by

$$f_d(\lambda) = \begin{cases} \rho_j & \text{if } \lambda = \pm\lambda_j, \\ 0 & \text{otherwise;} \end{cases}$$

and its *spectral distribution function* is given by

$$F(\lambda) = \int_{-\pi}^{\lambda} f_c(\omega) d\omega + \sum_{\omega \le \lambda} f_d(\omega), \qquad -\pi \le \lambda \le \pi.$$

The autocovariance is given in terms of the spectral distribution function by

$$\gamma(u) = \int_{-\pi}^{\pi} \exp(iu\lambda) dF(\lambda).$$

Note that f_c and f_d are symmetric about zero. If $p = 0$, then $\sum_u |\gamma(u)| < \infty$, $f_d = 0$, and the spectral distribution is absolutely continuous. We refer to $\pm\lambda_j$, $1 \le j \le p$, as the *atoms* of the spectral distribution and to ρ_j as the *mass* of the distribution at $\pm\lambda_j$. Note that if R_j equals $2\sqrt{\rho_j}$ in (8.1.1), then the time series is the mixed model discussed in Mackisack and Poskitt (1990).

Lspec is an adaptive methodology for estimating the spectral distribution for the series X_t. In particular, the log of the spectral density function is estimated with cubic splines and the line spectrum by a sum of Dirac delta functions. The estimation procedure to be described in Section 8.2 will be based upon *the periodogram*

$$I^{(T)}(\lambda) = (2\pi T)^{-1} \left| \sum_{t=0}^{T-1} \exp(-i\lambda t) X_t \right|^2, \qquad -\pi \le \lambda \le \pi,$$

corresponding to the realization $X_0, \ldots, X_{T-1}$ of the time series. If F is absolutely continuous, then the periodogram can be regarded as an (inconsistent) estimate of the spectral density function. In particular, if $\sum_u |u|\,|\gamma(u)| < \infty$ then

$$E[I^{(T)}(\lambda)] = f_c(\lambda) + (2\pi T)^{-1} \left[\frac{\sin T\lambda/2}{\sin \lambda/2} \right]^2 E(X_0)^2 + O(T^{-1}),$$

where the $O(T^{-1})$ is uniform in λ (Brillinger 1981).

It is convenient to refer to the function $f = f_c + \frac{T}{2\pi} f_d$ as the *mean function*. Under the assumptions that the atoms are all of the form $2\pi j/T$ for some integer j and that the time series is Gaussian, it can be shown that

$$I^{(T)}\left(\frac{2\pi j}{T}\right) = \left[f_c\left(\frac{2\pi j}{T}\right) + \frac{T}{2\pi} f_d\left(\frac{2\pi j}{T}\right)\right] W_j = f\left(\frac{2\pi j}{T}\right) W_j$$

for $1 \le j \le T/2$, where W_j has approximately the exponential distribution with mean one if $j < T/2$ and approximately the χ^2 distribution with one degree of freedom if T is even and $j = T/2$, and W_j, $1 \le j \le T/2$, are asymptotically independent; see Brillinger (1981, Theorem 5.2.6).

Since f_c and f_d are symmetric about zero, from now on we limit our attention to the interval $[0, \pi]$. Observe that if the indicated derivatives of f_c exist, then $f_c'(0)$, $f_c'''(0)$, $f_c'(\pi)$ and $f_c'''(\pi)$ all equal zero. Let $\delta_a(\lambda)$ equal one or zero according as $\lambda = a$ or $\lambda \ne a$. Set $\varphi = \log f$ and $\varphi_c = \log f_c$. Then $\varphi = \varphi_c + \varphi_d$, where $\varphi_d = \theta_1 \delta_{\lambda_1} + \cdots + \theta_p \delta_{\lambda_p}$ with $\theta_1, \ldots, \theta_p > 0$. Moreover, $f_d = (2\pi/T) f_c [\exp(\varphi_d) - 1]$. In Section 8.2 we will use cubic splines to obtain a finite-dimensional approximation to φ_c and hence to φ.

8.1.4 An Lspec model for the network data

In Figures 8.7 we show the Lspec estimates for the spectral density function and the line spectrum for the network data that was discussed in Section 8.1.1. We note that for both sets of data, the Lspec procedure includes an atom at $\frac{126}{36288} 2\pi$. Since the data summarizes 126 days, which equals 36288 five minute intervals, we conclude that Lspec indeed finds a discrete component of period one day in the spectral distribution. For the data that goes in the file server this component has about 6.6% of the mass of the spectral distribution, while for the data that comes out of the server this periodicity contributes about 6.8% of the mass of the spectral distribution. For both estimates Lspec also includes atoms at some harmonic components of the one day period: for the data that goes out of the file server there is one atom at frequency $\frac{252}{36288} 2\pi$ (a period of 0.5 day) with mass 1.4%, for the data that goes in the file server there are atoms at three different harmonic frequencies, with a combined mass of almost 3%. Since atoms at harmonic frequencies are included, we confirm that the periodicity is severely non-sinusoidal, something we also noted in Figures 8.4 and 8.6. Otherwise the spectrum is highly concentrated near the origin, which suggests strong non-periodic short time dependence of the data. The Lspec estimates of the spectral density for the data both going into and going out of the file-server use 16 knots, which explains the many details in the spectrum.

It is also interesting that for neither of the two data sets does the Lspec estimate include an atom at frequency $\frac{18}{36288} 2\pi$, corresponding to a period

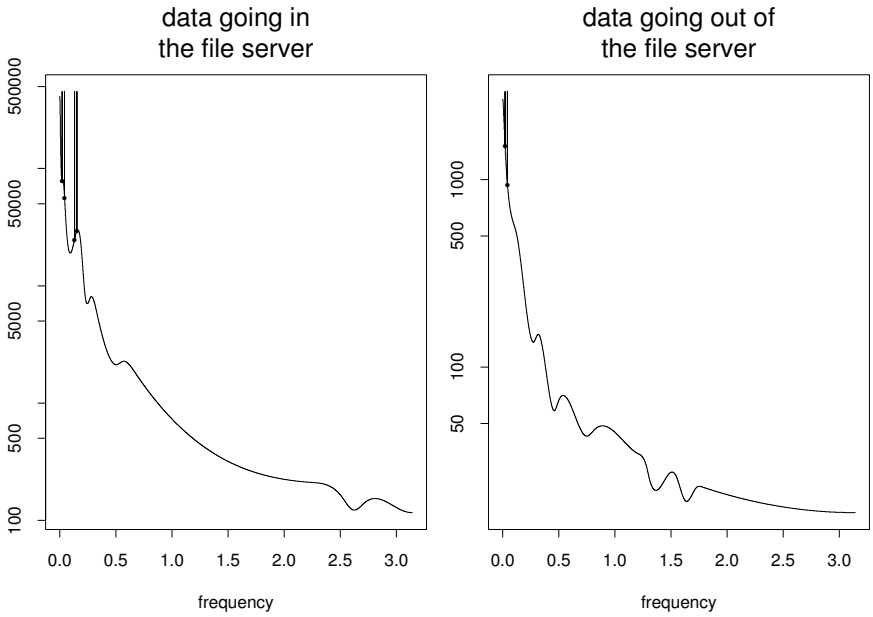

FIGURE 8.7. Lspec estimates of the spectral density functions and line function for the network data. Left: data going into the file server; right: data going out of the file server. Atoms are indicated by spikes starting at the height of the spectral density function.

of one week. In particular, Figure 8.3 suggests that there may be such a periodicity. We will get back to the issue of the possible weekly periodicity in Section 8.3. We also will discuss in that section what happens if a periodicity is not exactly of the form $2\pi j/T$.

If we would have analyzed the logarithm of the network data, which seems reasonable given the skewedness of the data, the mass of the spectral distribution that is explained by the atoms increases to about 15 to 20%. Since the possible periodicity with period one week disappears on a log-scale, we decided to focus on the data on its original scale.

8.2 The Lspec methodology

8.2.1 The Lspec model

The log of the spectral density function is modeled using cubic splines. Given the positive integer K_c and the sequence $t_1, \ldots, t_{K_c}$ of knots with $0 \leq t_1 < \cdots < t_{K_c} \leq \pi$, let $\mathbb{G}_c$ be the K_c-dimensional space of twice continuously differentiable functions g on $[0, \pi]$ such that the restriction of g to each of the intervals $[0, t_1], [t_1, t_2], \ldots, [t_{K_c-1}, t_{K_c}], [t_{K_c}, \pi]$ is a cubic polynomial, the first derivative of g is zero at 0 and π, the third derivative

of g is zero at 0 unless $t_1 = 0$, and the third derivative of g is zero at π unless $t_{K_c} = \pi$. (In particular, if $K_c = 1$, then $\mathbb{G}_c$ is the space of constant functions.) Note that the functions in $\mathbb{G}_c$ can be extended to splines on $(-\infty, \infty)$ that are symmetric about zero, periodic with period 2π, have a knot at zero if and only if $t_1 = 0$, and have a knot at π if and only if $t_{K_c} = \pi$. As we did for the Logspline basis (see Section 6.2.2) we can choose as a basis for $\mathbb{G}_c$ multiples of B-splines and a few additional basis functions that are nonzero at 0 and π. We omit the details.

Next, we describe the space that will be used indirectly to model the line spectrum. Given the nonnegative integer K_d and the increasing sequence $a_1, \ldots, a_{K_d}$ of members of $\{2\pi j/T : 1 \le j \le T/2\}$, let $\mathbb{G}_d$ be the K_d-dimensional space of nonnegative functions g on $[0, \pi]$ such that $g = 0$ except at $a_1, \ldots, a_{K_d}$. Set $B_{j+K_c}(\lambda) = \delta_{a_j}(\lambda)$ for $1 \le j \le K_d$. Then $B_{K_c+1}, \ldots, B_K$ form a basis of $\mathbb{G}_d$, where $K = K_c + K_d$.

Let $\mathbb{G}$ be the space spanned by $B_1, \ldots, B_K$. Set

$$\varphi_c(\cdot; \boldsymbol{\theta}_c) = \theta_1 B_1 + \cdots + \theta_{K_c} B_{K_c}$$

for $\boldsymbol{\theta}_c = [\theta_1, \ldots, \theta_{K_c}]^{\mathrm{T}} \in \mathbb{R}^{K_c}$,

$$\varphi_d(\cdot; \boldsymbol{\theta}_d) = \theta_{K_c+1} B_{K_c+1} + \cdots + \theta_K B_K$$

for $\boldsymbol{\theta}_d = [\theta_{K_c+1}, \ldots, \theta_K]^{\mathrm{T}}$ with $\theta_{K_c+1}, \ldots, \theta_K \ge 0$, and

$$\varphi(\cdot; \boldsymbol{\theta}) = \varphi_c(\cdot; \boldsymbol{\theta}_c) + \varphi_d(\cdot; \boldsymbol{\theta}_d)$$

for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^t$. We use $\varphi_c(\cdot; \boldsymbol{\theta}_c)$ to model the logarithm of the spectral density function and $\varphi(\cdot; \boldsymbol{\theta})$ to model the logarithm of the mean function. Thus, the spectral density function corresponding to $\boldsymbol{\theta}$ is given by $f_c(\cdot; \boldsymbol{\theta}_c) = \exp \varphi_c(\cdot; \boldsymbol{\theta}_c)$, the mean function is given by $f(\cdot; \boldsymbol{\theta}) = \exp \varphi(\cdot; \boldsymbol{\theta})$, and the line spectrum is given by

$$f_d(\cdot; \boldsymbol{\theta}_c) = \frac{2\pi}{T} f_c(\cdot; \boldsymbol{\theta}_c)[\exp(\varphi_d(\cdot; \boldsymbol{\theta}_d)) - 1].$$

Let $Y = f(\lambda; \boldsymbol{\theta})W$, where W has the exponential distribution with mean one when $0 < \lambda < \pi$ and it has the χ^2 distribution with one degree of freedom when $\lambda = \pi$. The log-likelihood corresponding to the observed value y of Y is given by

$$\psi(y, \lambda; \boldsymbol{\theta}) = \left(\frac{\delta_\pi(\lambda)}{2} - 1 \right) \left[\sum_{j=1}^{K} \theta_j B_j(\lambda) + y \exp \left(- \sum_{j=1}^{K} \theta_j B_j(\lambda) \right) \right]$$

for $0 < \lambda \le \pi$ and $y \ge 0$, where we have ignored a term that does not depend on $\boldsymbol{\theta}$. Observe that

$$\frac{\partial}{\partial \theta_k} \psi(y, \lambda; \boldsymbol{\theta}) = \left(\frac{\delta_\pi(\lambda)}{2} - 1 \right) B_k(\lambda) \left[1 - y \exp \left(- \sum_{j=1}^{K} \theta_j B_j(\lambda) \right) \right]$$

for $1 \le k \le K$, $0 < \lambda \le \pi$ and $y \ge 0$. Observe also that

$$\frac{\partial^2}{\partial\theta_k \partial\theta_l}\psi(y,\lambda;\boldsymbol{\theta}) = \left(\frac{\delta_\pi(\lambda)}{2} - 1\right) y B_k(\lambda) B_l(\lambda) \exp\left(-\sum_{j=1}^{K} \theta_j B_j(\lambda)\right)$$

for $1 \le k, l \le K$, $0 < \lambda \le \pi$ and $y \ge 0$. It follows from the last result that $\psi(y,\lambda;\cdot)$ is a concave function for $y \ge 0$ and $0 < \lambda \le \pi$.

Let $X_0, \ldots, X_{T-1}$ be a realization of length T of the time series. For $1 \le j \le T/2$, let

$$I_j = (2\pi T)^{-1} \left| \sum_{t=0}^{T-1} \exp(-i2\pi jt/T) X_t \right|^2$$

be the value of the corresponding periodogram at the (angular) frequency $\lambda = 2\pi j/T$. The (approximate) log-likelihood function corresponding to the periodogram and the K-parameter model for the logarithm of the mean function is given by

$$\ell(\boldsymbol{\theta}) = \sum_j \psi\left(I_j, \frac{2\pi j}{T}; \boldsymbol{\theta}\right), \qquad \boldsymbol{\theta} \in \mathbb{R}^K \text{ with } \theta_{K_c+1}, \ldots, \theta_K \ge 0,$$

which is equivalent to the Whittle likelihood (Whittle 1961). The maximum likelihood estimate $\widehat{\boldsymbol{\theta}} = [\hat\theta_1, \ldots, \hat\theta_K]^{\mathrm{T}}$ is given as usual by

$$\ell(\widehat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \mathbb{R}^K} \ell(\boldsymbol{\theta}),$$

the log-likelihood of the model is given by $\hat\ell = \ell(\widehat{\boldsymbol{\theta}})$, and the maximum likelihood estimate of the mean function is given by $\hat f(\lambda) = f(\lambda; \widehat{\boldsymbol{\theta}})$. Similarly, the maximum likelihood estimates of the spectral density function and line spectrum are given by $\hat f_c(\cdot) = f_c(\cdot; \widehat{\boldsymbol{\theta}}_c)$ and $\hat f_d(\cdot) = f_d(\cdot, \widehat{\boldsymbol{\theta}}_d)$, where $\widehat{\boldsymbol{\theta}}_c = (\hat\theta_1, \ldots, \hat\theta_{K_c})^T$ and $\widehat{\boldsymbol{\theta}}_d = (\hat\theta_{K_c+1}, \ldots, \hat\theta_K)^T$. For fixed $\mathbb{G}_c$ and $\mathbb{G}_d$ the maximum likelihood estimate can be found using a Newton–Raphson algorithm.

8.2.2 Model selection for Lspec models

Model selection for Lspec is carried out using a stepwise algorithm similar to those discussed in earlier chapters. However, the selection of knots and, in particular, atoms for Lspec is a much more ill-posed process than for many of the other procedures discussed. Thus, rather than fitting one sequence of models, we usually first go through one sequence of stepwise knot and atom addition and deletion, selecting the best Lspec model using AIC. We then use this best Lspec model that was obtained during the first

sequence of stepwise model selection as the initial model for a second cycle of stepwise model selection. We continue these cycles of a combination of stepwise addition followed by stepwise deletion until either the optimal model does not change or we reach a maximum number of cycles. In our experience, when the true underlying spectral distribution is mixed, one cycle of stepwise addition and deletion might not be enough to find the proper balance between the continuous component and the discrete component of the estimated spectral distribution. We have found in such situations that a two-cycle or multi-cycle procedure is more likely to find a good model.

During the first cycle, we initially fit the constant model. The location of the one knot in this model is of no importance. For simplicity we take $t_1 = 0$. After the constant model is fit, we add one knot according to the procedure described below. Then we remove the knot at zero and are left with a constant model having one knot. We continue by successively adding a knot or atom at each step. When searching for the location of the knot or atom in adding a basis function, we compute the Rao statistic for the addition of a knot for every frequency $\lambda = 2\pi j/T$, $1 \le j \le T/2$, such that $|\lambda - t_k| \ge 2(2\pi/T)$ for all knots t_k already in the model, and we compute the Rao statistic for the addition of an atom for every frequency $\lambda = 2\pi j/T$, $1 \le j \le T/2$, such that $\lambda \ne a_k$ for all atoms a_k already in the model. Since it is possible to compute Rao statistics for all candidate knots in $O(K_c M)$ time (see Kooperberg, Stone, and Truong 1995c), we do not need any heuristics to limit the number of candidate knots as was the case for several other of the methodologies we discussed. After we have reached a largest prespecified dimension, we proceed with a cycle of stepwise deletion. Among the fitted models we select the one that minimizes AIC (3.2.29) with default penalty parameter $a = \log n = \log(T/2)$.

To keep Lspec from selecting a model with too many atoms, we typically do not allow atoms to enter the model during the first cycle of stepwise knot addition and deletion, and we require all atoms to have a minimum mass (see Kooperberg, Stone, and Truong 1995c for details).

8.3 Further analysis of the network data

The most peculiar aspects of the analysis of the network data is probably the absence of an atom for the period of one week for the data going into the file server. To examine this further in Figure 8.8 we show on a square-root scale the first 300 entries of the raw periodogram

$$I^{(T)}\left(\frac{2\pi j}{T}\right),$$

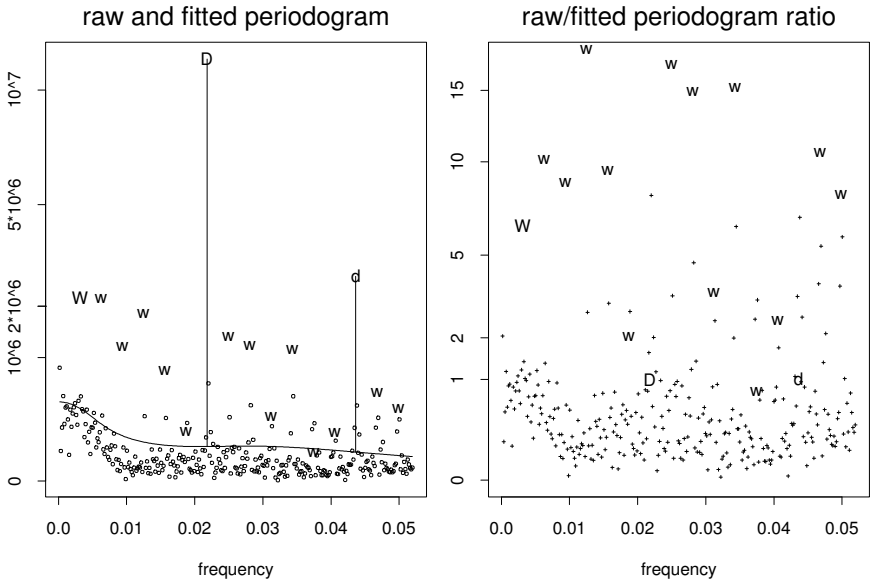

FIGURE 8.8. Left: raw periodogram (open symbols and letters) and fitted periodogram by Lspec (lines); right: ratio of raw periodogram and fitted periodogram.

the fitted values,

$$\widehat{I}^{(T)}\left(\frac{2\pi j}{T}\right) = \widehat{f}_c\left(\frac{2\pi j}{T}\right) + \frac{T}{2\pi}\widehat{f}_d\left(\frac{2\pi j}{T}\right),$$

and the ratio $I^{(T)}/\widehat{I}^{(T)}$. In the raw periodogram and in the ratios, we indicate with "W" the frequency corresponding to a period of one week, with "D" the frequency corresponding to period one day, with "w" the frequencies corresponding to harmonic frequencies of period one week that are not harmonic frequencies of period one day, and with "d" harmonic frequencies of period one day. As can be seen from this figure, the harmonic frequencies of periods one day and one week do indeed have the largest raw periodogram values, but the harmonic frequencies of one week do not particularly stand out. In particular, the ratio $I^{(T)}/\widehat{I}^{(T)}$ is larger for some of the harmonic frequencies corresponding to period one week than to the exact frequency corresponding to period one week, all suggesting that the within day pattern is very badly modeled by sinusoids.

Since Figure 8.8 suggested that there may be a small atom at a frequency corresponding to period one week, we also ran Lspec on the data going into the file-server, where we specified atoms at frequencies $(2\pi 18)/T$ and $(2\pi 126)/T$ corresponding to periods of one week and one day respectively. When we did this, Lspec did indeed retain the atom at $(2\pi 18)/T$ (with a mass of about 1% of the mass of the total spectral distribution), The graph of the spectral density looked very similar to the one shown in Figure 8.7.

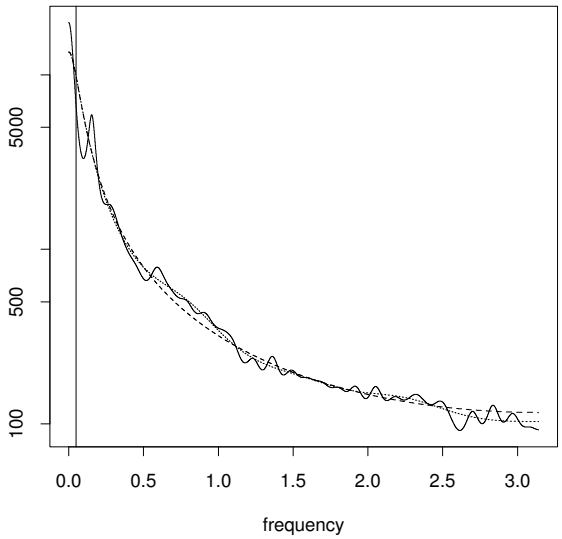

FIGURE 8.9. Autoregressive spectra for the data going into the file server: solid = order 58 (AIC), dotted = order 10, dashed = order 1. The solid vertical line corresponds to a period of one day.

The BIC value of the Lspec model with the atom was 266,469.1, which, interestingly, is much worse than the BIC value of the model without this atom (266,316.6).

An autoregressive model, using either BIC or AIC to select the order does not show any sign of an atom at frequency $(2\pi 126)/T$. Both methods select very high order models In Figure 8.9 we show the autoregressive spectrum for the AR model of order 58, which was selected by AIC. Interestingly, an AR(1) model (dashed) can already model the main shape of this spectrum, while an AR(10) model (dotted) follows the more complicated spectrum very closely, without the small bumps.

In the network data, a period of one day corresponds to a frequency of $(2\pi j)/T$ for $j = 126$, i.e. j is integer. Since Lspec can only put a mass at frequencies that are integer multiples of $(2\pi)/T$, it is of interest to examine what happens if there is a period in the data that corresponds to a frequency that is not of the specified form. To this end we also fitted an Lspec model to the same network data for the period August 25, 1998, 9:20pm through December 31, 1998, 5am. For the data going into the file-server a periodicity of one day corresponds to a frequency of $(2\pi 127.323)/T$. As it turned out, Lspec fitted a model with 2 atoms at frequencies corresponding to periods of 205.8 minutes and of 180.1 minute, which is approximately 1/7th ad 1/6th day, but no atom at a frequency close to one day. The Lspec model for the data going out of the file server, as well as the Lspec models for the logarithm of both series all included atoms with frequencies of $(2\pi 127)/T$,

corresponding to a period of one day and 43.9 minutes, which is the closest to a period of one day that Lspec can fit a model.

8.4 Extensions

Clearly there are currently two main drawbacks of the Lspec methodology:

1. Lspec can only includes atoms at frequencies that are exactly of the form $(2\pi j)/T$, for j integer.

2. The model selection aspect of Lspec is much more ill-posed than that for many of the other methods discussed in this book, such as Logspline (Chapter 6) and Hare (Chapter 7). Only Triogram (Chapter 9) may come a little close in ill-posedness, and even there the problems are less severe. In short, depending on the starting model some atoms may or may not end up in the model. An atom may or may not be in the model selected by Lspec depending on whether (another) atom was in the initial model or not. This is particular a problem for somewhat smaller atoms, and atoms for frequencies that are not of the form $(2\pi j)/T$ with j an integer.

To allow atoms at frequencies that are not of the form $(2\pi j)/T$, for j an integer, we can extend the approach in Section 8.2.1 by defining basis functions corresponding to a frequency λ. These basis functions are closely related to the Fourier transforms of sinusoids with frequency λ. Unfortunately, these basis functions would be nonzero for all frequencies of the form $(2\pi j)/T$, severely increasing the computational burden of Lspec. In addition, allowing for these additional atoms would undoubtedly make the worsen the model selection problem mentioned above.

To alleviate the model selection problem it may be of interest to study a simulated annealing or Markov chain Monte Carlo approach to Lspec, similar to what is done in Chapter 10 for Logspline and Triogram.

8.5 Notes

Literature

Cogburn and Davis (1974) uses smoothing splines to approximate the log periodogram. Rather than using the Whittle (1961) likelihood (which is used in Lspec) they used least squares regression to carry out the actual fitting. The smoothing splines are forced to be periodic. Wahba (1980) provided an automatic procedure to choose the smoothing parameter λ for the procedure of Cogburn and Davis (1974). Pawitan and O'Sullivan (1994) also used smoothing splines, but they fit the spline to the log-periodogram

using the same Whittle likelihood that is used for Lspec, however, their procedure does not allow for atoms, as it assumes the spectral density to be continuous.

Software

A program for implementing Lspec as described in this chapter has been written in C and an interface based on R has been developed. This code is currently available from `CRAN`. Several others have ported this S-Plus and C code for easy installation on other platforms and under the R language. See Kooperberg's website

`http://bear.fhcrc.org/~clk/soft.html`

for current links.

It should be noted that an approximate, nonadaptive version of Lspec is readily available in many statistical packages. Specifically, set `freq` $= 2\pi j/T$ and `period` $= I^{(T)}(2\pi j/T)$ for $1 \le j \le T/2$ and let `atom1` be a vector of the form $(0, \ldots, 0, 1, 0, \ldots, 0)^t$ of length $\lfloor T/2 \rfloor$ with the 1 in the jth position. Then the S-Plus command

```
fixed <-  glm(period ~ bs(freq,k) + atom1,
                family = quasi(log, "mu^2"))
```

fits a B-spline with k degrees of freedom and an atom at $2\pi j/T$ to the spectral distribution. However, the location and number of the knots and atoms are not optimized, and the first and third derivatives of the spectral density at 0 and π are not constrained to equal 0.

Theory

Under suitable conditions, Kooperberg, Stone, and Truong (1995d) obtained the L_2 rate of convergence for a nonadaptive version of Logspline spectral density estimation. This result lends theoretical support to Lspec.