Notes on Non-parametric estimation

To tie up the discussion on the topic of estimation, more specifically spectral estimation, we derive here some tighter properties related to the direct use of the Fourier Transform, as contrasted to the properties of parametric methods such as the Levinson-Schur algorithm, where parameters of a model are estimated. We shall obtain a better insight in how well the Fourier Transform on discrete stochastic time series is capable to estimate the Power Spectral Density Function (PSDF). In this section we follow the book of Stoica and Moses as mentioned, because it gives a more succinct description of the topic than can be found in the book of Porat. The discussion also gives a nice introduction to the topic of 'windows' used to preprocess the data so that the quality of the spectral estimation is enhanced to some extent.

Periodogram vs. Correlogram

Traditionally, two approaches to estimate the PSDF have been defined. We assume that we dispose of a chunk of data consisting of regular samples and that our goal is to derive the PSDF from the given data (much as shown in the section on the Schur algorithm). One method is called the periodogram. It consists simply in taking the discrete Fourier transform of the data vector, and using its absolute value square as the value for samples of the PSDF. The other method is to start from the definition of the PSDF as the Fourier Transform of the covariance lags $r[n]$. Since we do not know these lags, because they are in fact expectations of a stochastic variable, we have to estimate them from the data we have available. This can be done in a variety of ways, two of which we shall consider in detail in the following sheets.

Estimates of the covariance lags

One method is known as the 'standard unbiased method', and it consists in taking a correct time average of the product of the given series with a shifted (and complex conjugated) version of it, as shown in the formula. If the shift is over $k$ data points, then we are actually missing $k$ terms in the series, so that a correct average has to be taken over $N-k$ terms. Another method is called 'standard biased' and it consists in taking the same sum but averaging always over the same number $N-1$. There are conceivably other possibilities, we consider here only these two, because they are the most important.

Property 1

It turns out that the periodogram gives the same result as the standard biased estimate. That is a pleasant result because it allows us to use a straight FFT to compute the PSDF in the standard biased version. The proof is simple by direct, brute force computation, we skip it!

Discussion

Another nice consequence of property 1 is that we can study the statistical properties of the periodogram by analyzing the standard biased correlogram. We shall do that now to some extent, and compare the results with what is obtained for the standard unbiased method. Surprisingly, the standard biased method can be said to be generally as good as
the standard unbiased. To indicate that there are problems with non-parametric estimation, we make a quick qualitative analysis of the correlogram method. There are two competing mechanisms that influence accuracy. One is the 'law of large numbers' which says that an average quantity over a large number of otherwise equal and independent experiments converges in probability to the expectation (the ensemble average), whereby the rate of converge is roughly proportional to $1/\sqrt{N}$ - in other words, the terms in the correlogram have relative accuracy of that order. The other is that to obtain the correlogram we actually sum $N$ terms. If all these terms were equal in magnitude then the error propagation in the sum would have the character of a random walk, and one could expect a worst case behavior that produces an error on the average as big as the error on a single sample! Luckily, that is not the case in practice, mainly because the covariance lags decay fast enough so that their contributions become negligible in the sum. In the case of the standard biased method, the situation is even more favorable because, as we shall see, the effect of this method is to apply a triangular window on the lags, making the contribution of larger lags even smaller. Be that as it may, a thorough investigation of the statistical properties of the method is generally needed in practice (it is given in the book mentioned). It goes, however, beyond the present treatment, we refer to the literature for further information.

However, we can get a number of elementary results just by studying the expectation of the correlograms. When we use time estimates for the covariance lags, then these quantities are samples of actual stochastic variables in their own right (they are sums of products of samples of the original stochastic variables), and we can perform statistical analysis on them. In particular, we can calculate their expectation (we could also calculate their variance, and that would give us good estimates of accuracy, but as announced, we shall skip these more sophisticated computations, see the literature if you need more information). Using the direct expressions for the estimates of the covariance lags, we quickly find the expectations needed. Aside from 'statistical noise', these expectations should give us a clear indication of the properties of our PSDF estimates.

**Properties of the periodogram**

We now embark on the computation of the properties of the periodogram via its expectation (and we skip the accuracy effects on the numerical results obtained - in other words, we do as if we have enough statistical accuracy). Because of the bias in the sum, we see that the covariance lags in the standard biased correlogram are multiplied with a coefficient $1 - \frac{|n|}{N}$, in other words by a triangular window (as shown on the slide). What is the effect on the PSDF? To evaluate that we use the reverse convolution theorem for the Fourier Transform, which says that the Fourier Transform of the product of two series is actually given by the convolution on the unit circle of the corresponding Fourier Transforms. This is indicated on the slide, where the Fourier Transform of the triangular window is denoted by $W^d(\theta)$. The distortion on the original spectrum (abstraction made of statistical errors) is hence given by the convolution with the spectrum of the window. Hence, we have to evaluate this windowing spectrum and study its properties! It is easy to write down the formula for the spectrum: $NW_b^d(\theta) = \sum_{n=-N+1}^{N-1} (N - |n|)e^{-j\theta n}$. How to
evaluate this sum? The next slide shows a little trick. In abscis we put all the different values of \( e^{-jm\theta} \) we have to use, ordered by \( n \), while in ordinate we just put dots equal to the number of times we have to use that value, in such a way that a nice pattern of terms to be summed appears in the form of a parallelogram. For example, we see that for \( n=N-1 \) we have just one value of the corresponding \( e^{-j(N-1)\theta} \) in the sum while for \( n=0 \), we have \( N \) values. If we now sum the dots along SW-NE diagonal we find that, except for a common factor, all diagonals sum to the same value, and the overall value is found by multiplying that quantity with the sum of the respective coefficients, which again sum to the complex conjugate of the same value. The triangular window, also known as the Bartlett window, hence has the Fourier Transform as indicated.

**The Bartlett window**

The resulting spectrum for the Bartlett window has the following properties:

1. **it is always positive**;
2. **it looks like a smeared out Dirac Impulse.** In the neighborhood of \( \theta = 0 \) it looks like a function, i.e. it has a pretty sharp peak and reaches zero tangentially at \( \theta = \pm \frac{\pi}{N} \). One can safely say that its half amplitude pulse width is approximately \( \frac{\pi}{N} \) as well. Because of the quadratic nature of the response, the 'smearing' of a signal localized at a given frequency to other frequencies quickly goes to zero, much quicker than what is the case with the Dirichlet window or rectangular window valid for the unbiased case (see the resulting graphs to have an idea).

**Unbiased estimate: the Dirichlet window**

A similar, even simpler analysis gives the spectral distortion for the standard unbiased correlogram. The Fourier transform of the rectangular window valid for this case is of the form \( \frac{\sin x}{x} \) (without a square) for \( x = (N - \frac{1}{2})\theta \), with a half bandwidth that is somewhat smaller than the preceding, but much stronger smearing.

**Other windows**

We leave the discussion at this point, just indicate that there are a number of other possibilities. One can use different kinds of windows, not letting them extend to \( N \), and use different kinds of estimates of the data. There is a large literature on this topic!