Chapter 8

Spectral Representations

Prerequisites

- Knowledge of complex numbers.
- Have some idea of what the covariance of a complex random variable (we do define it below).
- Some idea of a Fourier transform (a review is given in Section A.3).

Objectives

- Know the definition of the spectral density.
- The spectral density is always non-negative and this is a way of checking that a sequence is actually non-negative definite (is a autocovariance).
- The DFT of a second order stationary time series is almost uncorrelated.
- The spectral density of an ARMA time series, and how the roots of the characteristic polynomial of an AR may influence the spectral density function.
- There is no need to understand the proofs of either Bochner’s (generalised) theorem or the spectral representation theorem, just know what these theorems are. However, you should know the proof of Bochner’s theorem in the simple case that $\sum_r |rc(r)| < \infty$. 

227
8.1 How we have used Fourier transforms so far

We recall in Section 1.2.4 that we considered models of the form

$$X_t = A \cos (\omega t) + B \sin (\omega t) + \varepsilon_t \quad t = 1, \ldots, n. \tag{8.1}$$

where $\varepsilon_t$ are iid random variables with mean zero and variance $\sigma^2$ and $\omega$ is unknown. We estimated the frequency $\omega$ by taking the Fourier transform $J_n(\omega) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t e^{it\omega}$ and using as an estimator of $\omega$, the value which maximised $|J_n(\omega)|^2$. As the sample size grows the peak (which corresponds the frequency estimator) grows in size. Besides the fact that this corresponds to the least squares estimator of $\omega$, we note that

$$J_n(\omega_k) = \frac{1}{2\pi n} \sum_{t=1}^{n} X_t \exp(i t \omega_k)$$

$$= \frac{1}{2\pi n} \sum_{t=1}^{n} \mu(t/n) \exp(i t \omega_k) + \frac{1}{2\pi n} \sum_{t=1}^{n} \varepsilon_t \exp(i t \omega_k) \tag{8.2}$$

where $\omega_k = \frac{2\pi k}{n}$, is an estimator the the Fourier transform of the deterministic mean at frequency $k$. In the case that the mean is simply the sin function, there is only one frequency which is non-zero. A plot of one realization ($n = 128$), periodogram of the realization, periodogram of the iid noise and periodogram of the sin function is given in Figure 8.1. Take careful note of the scale (y-axis), observe that the periodogram of the sin function dominates the the periodogram of the noise (magnitudes larger). We can understand why from (8.2), where the asymptotic rates are given and we see that the periodogram of the deterministic signal is estimating $n \times$ Fourier coefficient, whereas the periodogram of the noise is $O_p(1)$. However, this is an asymptotic result, for small samples sizes you may not see such a big difference between deterministic mean and the noise. Next look at the periodogram of the noise we see that it is very erratic (we will show later that this is because it is an inconsistent estimator of the spectral density function), however, despite the erraticness, the amount of variation overall frequencies seems to be same (there is just one large peak - which could be explained by the randomness of the periodogram).

Returning again to Section 1.2.4, we now consider the case that the sin function has been
corrupted by colored noise, which follows an AR(2) model

\[ \epsilon_t = 1.5\epsilon_{t-1} - 0.75\epsilon_{t-2} + \epsilon_t. \]  \hspace{1cm} (8.3)

A realisation and the corresponding periodograms are given in Figure 8.2. The results are different to the iid case. The peak in the periodogram no longer corresponds to the period of the sin function. From the periodogram of the just the AR(2) process we observe that it erratic, just as in the iid case, however, there appears to be varying degrees of variation over the frequencies (though this is not so obvious in this plot). We recall from Chapters 2 and 3, that the AR(2) process has a pseudo-period, which means the periodogram of the colored noise will have pronounced peaks which correspond to the frequencies around the pseudo-period. It is these pseudo-periods which are dominating the periodogram, which is giving a peak at frequency that does not correspond to the sin function. However, asymptotically the rates given in (8.2) still hold in this case too. In other words, for large enough sample sizes the DFT of the signal should dominate the noise. To see that this is the case, we increase the sample size to \( n = 1024 \), a realisation is given in Figure 8.3.

We see that the period corresponding the sin function dominates the periodogram. Studying the periodogram of just the AR(2) noise we see that it is still erratic (despite the large sample size),

Figure 8.1: Top Left: Realisation of (1.5) \((2 \sin(\frac{2\pi t}{8}))\) with iid noise, Top Right: Periodogram of \( \sin + \text{noise} \). Bottom Left: Periodogram of just the noise. Bottom Right: Periodogram of just the \( \sin \) function.
but we also observe that the variability clearly changes over frequency.

Figure 8.2: Top Left: Realisation of (1.5) \((2 \sin(\frac{2\pi t}{8}))\) with AR(2) noise \((n = 128)\), Top Right: Periodogram. Bottom Left: Periodogram of just the AR(2) noise. Bottom Right: Periodogram of the sin function.

Figure 8.3: Top Left: Realisation of (1.5) \((2 \sin(\frac{2\pi t}{8}))\) with AR(2) noise \((n = 1024)\), Top Right: Periodogram. Bottom Left: Periodogram of just the AR(2) noise. Bottom Right: Periodogram of the sin function.

From now on we focus on the constant mean stationary time series (eg. iid noise and the AR(2))
was first introduced in in Section 1.6. We recall that given an autoregressive process

\[ J_n(\omega_k) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \exp(it \omega_k). \]  

(8.4)

This is simply a (linear) transformation of the data, thus it easily reversible by taking the inverse DFT

\[ X_t = \frac{\sqrt{2\pi}}{\sqrt{n}} \sum_{t=1}^{n} J_n(\omega_k) \exp(-it \omega_k). \]  

(8.5)

Therefore, just as one often analyzes the log transform of data (which is also an invertible transform), one can analyze a time series through its DFT.

In Figure 8.4 we give plots of the periodogram of an iid sequence and AR(2) process defined in equation (8.3). We recall from Chapter 3, that the periodogram is an inconsistent estimator of the spectral density function \( f(\omega) = (2\pi)^{-1} \sum_{r=-\infty}^{\infty} c(r) \exp(ir \omega) \) and a plot of the spectral density function corresponding to the iid and AR(2) process defined in (8.3). We will show later that by inconsistent estimator we mean that \( \text{E}[|J_n(\omega_k)|^2] = f(\omega_k) + O(n^{-1}) \) but \( \text{var}[|J_n(\omega_k)|^2] \to 0 \) as \( n \to \infty \). this explains why the general ‘shape’ of \(|J_n(\omega_k)|^2\) looks like \( f(\omega_k) \) but \(|J_n(\omega_k)|^2\) is extremely erratic and variable.

![Figure 8.4: Left: Periodogram of iid noise. Right: Periodogram of AR(2) process.](image)

Remark 8.1.1 (Properties of the spectral density function) The spectral density function was first introduced in in Section 1.6. We recall that given an autoregressive process \( \{c(k)\} \), the
Figure 8.5: Left: Spectral density of iid noise. Right: Spectral density of AR(2), note that the interval [0, 1] corresponds to [0, 2π] in Figure 8.5.

The spectral density is defined as

\[ f(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c(r) \exp(2\pi ir). \]

And visa versa, given the spectral density we can recover the autocovariance via the inverse transform \( c(r) = \int_{0}^{2\pi} f(\omega) \exp(-2\pi ir\omega) d\omega. \) We recall from Section 1.6 that the spectral density function can be used to construct a valid autocovariance function since only a sequence whose Fourier transform is real and positive can be positive definite.

In Section 5.4 we used the spectral density function to define conditions under which the variance covariance matrix of a stationary time series had minimum and maximum eigenvalues. Now from the discussion above we observe that the variance of the DFT is approximately the spectral density function (note that for this reason the spectral density is sometimes called the power spectrum).

We now collect some of the above observations, to summarize some of the basic properties of the DFT:

(i) We note that \( J_n(\omega_k) = J_n(\omega_{n-k}) \), therefore, all the information on the time series is contained in the first \( n/2 \) frequencies \( \{J_n(\omega_k); k = 1, \ldots, n/2\} \).

(ii) If the time series \( E[X_t] = \mu \) and \( k \neq 0 \) then

\[ E[J_n(\omega_k)] = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mu \exp(it\omega_k) = 0. \]
If \( k = 0 \) then

\[
E[J_n(\omega_0)] = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mu = \sqrt{n}\mu.
\]

In other words, the mean of the DFT (at non-zero frequencies) is zero regardless of whether the time series has a zero mean (it just needs to have a constant mean).

(iii) However, unlike the original stationary time series, we observe that the variance of the DFT depends on frequency (unless it is a white noise process) and that for \( k \neq 0 \),

\[
\operatorname{var}[J_n(\omega_k)] = E|[J_n(\omega_k)]^2| = f(\omega_k) + O(n^{-1}).
\]

The focus of this chapter will be on properties of the spectral density function (proving some of the results we stated previously) and on the so called Cramer representation (or spectral representation) of a second order stationary time series. However, before we go into these results (and proofs) we give one final reason why the analysis of a time series is frequently done by transforming to the frequency domain via the DFT. Above we showed that there is a one-to-one correspondence between the DFT and the original time series, below we show that the DFT almost decorrelates the stationary time series. In other words, one of the main advantages of working within the frequency domain is that we have transformed a correlated time series into something that it almost uncorrelated (this also happens to be a heuristic reason behind the spectral representation theorem).

8.2 The ‘near’ uncorrelatedness of the Discrete Fourier Transform

Let \( X_n = \{X_t; t = 1, \ldots, n\} \) and \( \Sigma_n = \operatorname{var}[X_n] \). It is clear that \( \Sigma_n^{-1/2}X_n \) is an uncorrelated sequence. This means to formally decorrelate \( X_n \) we need to know \( \Sigma_n^{-1/2} \). However, if \( X_t \) is a second order stationary time series, something curiously, remarkable happens. The DFT, almost uncorrelates the \( X_n \). The implication of this is extremely useful in time series, and we shall be using this transform in estimation in Chapter 9.

We start by defining the Fourier transform of \( \{X_t\}_{t=1}^{n} \) as

\[
J_n(\omega_k) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \exp(ik\frac{2\pi t}{n})
\]
where the frequences $\omega_k = 2\pi k/n$ are often called the fundamental, Fourier frequencies.

**Lemma 8.2.1** Suppose $\{X_t\}$ is a second order stationary time series, where $\sum_r |rc(r)| < \infty$. Then we have

$$\text{cov}(J_n\left(\frac{2\pi k_1}{n}\right), J_n\left(\frac{2\pi k_2}{n}\right)) = \begin{cases} f\left(\frac{2\pi k}{n}\right) + O\left(\frac{1}{n}\right) & k_1 = k_2 \\ O\left(\frac{1}{n}\right) & 1 \leq k_1 \neq k_2 \leq n/2 \end{cases}$$

where $f(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c(r) \exp(ir\omega)$. If one wants to consider the real and imaginary parts of $J_n(\omega_k)$ then

$$\text{cov}(J_n,C\left(\frac{2\pi k_1}{n}\right), J_n,C\left(\frac{2\pi k_2}{n}\right)) = \begin{cases} f\left(\frac{2\pi k}{n}\right) + O\left(\frac{1}{n}\right) & k_1 = k_2 \\ O\left(\frac{1}{n}\right) & 1 \leq k_1 \neq k_2 \leq n/2 \end{cases}$$

and $\text{cov}[J_n,C\left(\frac{2\pi k_1}{n}\right), J_n,S\left(\frac{2\pi k_2}{n}\right)] = O(n^{-1})$ for $1 \leq k_1, k_2 \leq n/2$, where

$$J_n,C(\omega_k) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \cos(t\omega_k), \quad J_n,S(\omega_k) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \sin(t\omega_k).$$

In the sections below we give two proofs for the same result.

We note that the principle reason behind both proofs is that

$$\sum_{t=1}^{n} \exp\left(it \frac{2\pi j}{n}\right) = \begin{cases} 0 & j \neq n\mathbb{Z} \\ n & j \in \mathbb{Z} \end{cases} \quad (8.6)$$

### 8.2.1 ‘Seeing’ the decorrelation in practice

We evaluate the DFT using the following piece of code (note that we do not standardize by $\sqrt{2\pi}$)

```r
dft <- function(x){
    n=length(x)
    dft <- fft(x)/sqrt(n)
    return(dft)
}
```

234
We have shown above that \( \{J_n(\omega_k)\}_k \) are close to uncorrelated and have variance close to \( f(\omega_k) \). This means that the ratio \( J_n(\omega_k)/f(\omega_k)^{1/2} \) are close to uncorrelated with variance close to one. Let us treat

\[
Z_k = \frac{J_n(\omega_k)}{\sqrt{f(\omega_k)}},
\]

as the transformed random variables, noting that \( \{Z_k\} \) is complex, our aim is to show that the acf corresponding to \( \{Z_k\} \) is close to zero. Of course, in practice we do not know the spectral density function \( f \), therefore we estimate it using the piece of code (where \( \text{test} \) is the time series)

\[
k<-\text{kernel("daniell",6)}
\]

\[
temp2<-\text{spec.pgram(test,k, taper=0, log = "no")}$\text{spec}
\]

\[
n<-\text{length(temp2)}
\]

\[
temp3<-c(\text{temp2[c(1:n)]},\text{temp2[c(n:1)]})
\]

temp3 simply takes a local average of the periodogram about the frequency of interest (however it is worth noting that \( \text{spec.pgram} \) does not do precisely this, which can be a bit annoying). In Section 9.3 we explain why this is a consistent estimator of the spectral density function. Notice that we also double the length, because the estimator \( \text{temp2} \) only gives estimates in the interval \([0,\pi]\). Thus our estimate of \( \{Z_k\} \), which we denote as \( \widehat{Z}_k = J_n(\omega_k)/\widehat{f}_n(\omega_k)^{1/2} \) is

\[
temp1<-\text{dft(test)}; \text{ temp4<-temp1/sqrt(temp3)}
\]

We want to evaluate the covariance of \( \{\widehat{Z}_k\} \) over various lags

\[
\widehat{C}_n(r) = \frac{1}{n} \sum_{k=1}^{n} \widehat{Z}_k \overline{\widehat{Z}_{k+r}} = \frac{1}{n} \sum_{k=1}^{n} \frac{J_n(\omega_k)\overline{J_n(\omega_{k+r})}}{\sqrt{\widehat{f}_n(\omega_k)\overline{\widehat{f}_n(\omega_{k+r})}}}
\]

To speed up the evaluation, we use we can exploit the speed of the FFT, Fast Fourier Transform.

A plot of the AR(2) model

\[
\varepsilon_t = 1.5\varepsilon_{t-1} - 0.75\varepsilon_{t-2} + \varepsilon_t.
\]

together with the real and imaginary parts of its DFT autocovariance is given in Figure 8.6. We observe that most of the correlations lie between \([-1.96, 1.96]\) (which corresponds to the 2.5% limits
of a standard normal). Note that the 1.96 corresponds to the 2.5% limits, however this bound only holds if the time series is Gaussian. If the time series is non-Gaussian some corrections have to be made (see Dwivedi and Subba Rao (2011) and Jentsch and Subba Rao (2014)).

Figure 8.6: Top: Realization. Middle: Real and Imaginary of $\sqrt{n}\hat{C}_n(r)$ plotted against the 'lag' $r$. Bottom: QQplot of the real and imaginary $\sqrt{n}\hat{C}_n(r)$ against a standard normal.
Exercise 8.1  (a) Simulate an AR(2) process and run the above code using the sample size

(i) \( n = 64 \) (however use \( k \leftarrow \text{kernel("daniell",3)} \))

(ii) \( n = 128 \) (however use \( k \leftarrow \text{kernel("daniell",4)} \))

Does the ‘near decorrelation property’ hold when the sample size is very small. Explain your
answer by looking at the proof of the lemma.

(b) Simulate a piecewise stationary time series (this is a simple example of a nonstationary time
series) by stringing two stationary time series together. One example is

\[
\begin{align*}
\text{ts1} &= \text{arima.sim(list(order=c(2,0,0), ar = c(1.5, -0.75)), n=128)}; \\
\text{ts2} &= \text{arima.sim(list(order=c(1,0,0), ar = c(0.7)), n=128)} \\
\text{test} &= \text{c(ts1/sd(ts1),ts2/sd(ts2))}
\end{align*}
\]

Make a plot of this time series. Calculate the DFT covariance of this time series, what do
you observe in comparison to the stationary case?

8.2.2 Proof 1 of Lemma 8.2.1: By approximating Toeplitz with

Circulant matrices

Let \( \mathbf{X}_n' = (X_n, \ldots, X_1) \) and \( F_n \) be the Fourier transformation matrix \( (F_n)_{s,t} = n^{-1/2} \Omega_n^{(s-1)(t-1)} = n^{-1/2} \exp\left(\frac{2\pi i (s-1)(t-1)}{n}\right) \). It is clear that \( F_n \mathbf{X}_n = (J_n(\omega_0), \ldots, J_n(\omega_{n-1}))' \).

We now prove that \( F_n \mathbf{X}_n \) is almost an uncorrelated sequence.

The first proof will be based on approximating the symmetric Toeplitz variance matrix of \( \mathbf{X}_n \)
with a circulant matrix, which has well known eigen values and functions. We start by considering
the variance of \( F_n \mathbf{X}_n' \), \( \text{var}(F_n \mathbf{X}_n) = F_n \Sigma_n F_n' \), and our aim is to show that it is almost a diagonal.

We first recall that if \( \Sigma_n \) were a circulant matrix, then \( F_n \mathbf{X}_n \) would be uncorrelated since \( F_n \) is the
eigenmatrix of any circulant matrix. This is not the case. However, the upper right hand side and
the lower left hand side of \( \Sigma_n \) can approximated by circulant matrices - this is the trick in showing
the ‘near’ uncorrelatedness. Studying $\Sigma_n$

$$
\Sigma_n = \begin{pmatrix}
    c(0) & c(1) & c(2) & \ldots & c(n-1) \\
    c(1) & c(0) & c(1) & \ldots & c(n-2) \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    c(n-1) & c(n-2) & \vdots & c(1) & c(0)
\end{pmatrix}
$$

we observe that it can be written as the sum of two circulant matrices, plus some error, that we will bound. That is, we define the two circulant matrices

$$
C_{1n} = \begin{pmatrix}
    c(0) & c(1) & c(2) & \ldots & c(n-1) \\
    c(n-1) & c(0) & c(1) & \ldots & c(n-2) \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    c(1) & c(2) & \vdots & c(n-1) & c(0)
\end{pmatrix}
$$

and

$$
C_{2n} = \begin{pmatrix}
    0 & c(n-1) & c(n-2) & \ldots & c(1) \\
    c(1) & 0 & c(n-1) & \ldots & c(2) \\
    c(2) & c(1) & 0 & \ldots & c(3) \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    c(n-1) & c(n-2) & \vdots & c(1) & 0
\end{pmatrix}
$$

We observe that the upper right hand sides of $C_{1n}$ and $\Sigma_n$ match and the lower left and sides of $C_{2n}$ and $\Sigma_n$ match. As the above are circulant their eigenvector matrix is $F_n$ (note that $F_n^{-1} = F_n'$). Furthermore, the eigenvalues matrix of $C_{n1}$ is

$$
\text{diag} \left( \sum_{j=0}^{n-1} c(j), \sum_{j=0}^{n-1} c(j)\Omega_n, \ldots, \sum_{j=0}^{n-1} c(j)\Omega_n^{(t-1)j} \right),
$$
whereas the eigenvalue matrix of $C_{n2}$ is

$$
\begin{align*}
\text{diag} & \left( \sum_{j=1}^{n-1} c(j), \sum_{j=1}^{n-1} c(n-j)\Omega_n^j, \ldots, \sum_{j=1}^{n-1} c(n-j)\Omega_n^{(t-1)j} \right) \\
= & \text{diag} \left( \sum_{j=1}^{n-1} c(j), \sum_{j=1}^{n-1} c(j)\Omega_n^{-j}, \ldots, \sum_{j=1}^{n-1} c(j)\Omega_n^{-(t-1)j} \right),
\end{align*}
$$

More succinctly, the $k$th eigenvalues of $C_{n1}$ and $C_{n2}$ are $\lambda_k = \sum_{j=0}^{n-1} c(j)\Omega_n^{j(k-1)}$ and $\lambda_k = \sum_{j=1}^{n-1} c(j)\Omega_n^{-j(k-1)}$. Observe that $\lambda_k + \lambda_{k'} = \sum_{|j| \leq (n-1)} c(j)e^{2\pi j \omega} \approx f(\omega_j)$, thus the sum of these eigenvalues approximate the spectral density function.

We now show that under the condition $\sum_n |rc(r)| < \infty$ we have

$$
F_n \Sigma_n \overline{F}_n' - F_n (C_{n1} + C_{n2}) \overline{F}_n = O \left( \frac{1}{n} \right) I,
$$

where $I$ is a $n \times n$ matrix of ones. To show the above we consider the differences element by element. Since the upper right hand sides of $C_{n1}$ and $\Sigma_n$ match and the lower left and sides of $C_{n2}$ and $\Sigma_n$ match, the above difference is

$$
\begin{align*}
& \left| \left( F_n \Sigma_n \overline{F}_n' - F_n (C_{n1} + C_{n2}) \overline{F}_n \right)_{s,t} \right| \\
= & |\varepsilon_s \Sigma_n \overline{v'}_t - \varepsilon_s C_{n1} \overline{v'}_t - \varepsilon_s C_{n2} \overline{v'}_t| \\
\leq & \frac{2}{n} \sum_{r=1}^{n-1} |rc(r)| = O \left( \frac{1}{n} \right).
\end{align*}
$$

Thus we have shown (8.7). Therefore, since $F_n$ is the eigenvector matrix of $C_{n1}$ and $C_{n2}$, altogether we have

$$
F_n (C_{n1} + C_{n2}) \overline{F}_n = \text{diag} \left( f_n(0), f_n \left( \frac{2\pi}{n} \right), \ldots, f_n \left( \frac{2\pi(n-1)}{n} \right) \right),
$$

where $f_n(\omega) = \sum_{r=-n}^{(n-1)} c(r) \exp(i j \omega)$. Altogether this gives

$$
\text{var}(F_n \Sigma_n) = F_n \Sigma_n F_n = \begin{pmatrix}
  f_n(0) & 0 & \ldots & 0 & 0 \\
  0 & f_n \left( \frac{2\pi}{n} \right) & \ldots & 0 & 0 \\
  \vdots & \ddots & \ddots & \ddots & \vdots \\
  0 & \ldots & \ldots & 0 & f_n \left( \frac{2\pi(n-1)}{n} \right) \\
\end{pmatrix} + O \left( \frac{1}{n} \right) B.
$$

239
Finally, we note that since $\sum_r |rc(r)| < \infty$

$$|f_n(\omega) - f(\omega)| \leq \sum_{|r|>n} |c(r)| \leq \frac{1}{n} \sum_{|r|>n} |rc(r)| = O(n^{-1}),$$

which gives the required result.

**Remark 8.2.1** Note the eigenvalues of a matrix are often called the spectrum and that above calculation shows that spectrum of $\text{var}[X_n]$ is close to $f(\omega_n)$, which may be one reason why $f(\omega)$ is called the spectral density (the reason for density probably comes from the fact that $f$ is positive).

These ideas can also be used for inverting Toeplitz matrices (see Chen et al. (2006)).

### 8.2.3 Proof 2 of Lemma 8.2.1: Using brute force

A more hands on proof is to just calculate $\text{cov}(J_n(\frac{2\pi k_1}{n}), J_n(\frac{2\pi k_2}{n}))$. The important aspect of this proof is that if we can isolate the exponentials than we can use (8.6). It is this that gives rise to the near uncorrelatedness property. Remember also that $\exp(i\frac{2\pi}{n}jk) = \exp(ik\omega_k) = \exp(ik\omega_j)$, hence we can interchange between the two notations.

We note that $\text{cov}(A, B) = E(AB) - E(A)E(B)$, thus we have

$$\text{cov} \left( J_n \left( \frac{2\pi k_1}{n} \right), J_n \left( \frac{2\pi k_2}{n} \right) \right) = \frac{1}{n} \sum_{t, \tau = 1}^n \text{cov}(X_t, X_\tau) \exp \left( i(tk_1 - \tau k_2) \frac{2\pi}{n} \right)$$

Now change variables with $r = t - \tau$, this gives (for $0 \leq k_1, k_2 < n$)

$$\text{cov} \left( J_n \left( \frac{2\pi k_1}{n} \right), J_n \left( \frac{2\pi k_2}{n} \right) \right) = \frac{1}{n} \sum_{r = -(n-1)}^{n-1} c(r) \exp \left( -ir \frac{2\pi k_2}{n} \right) \sum_{t=1}^{n-|r|} \exp \left( \frac{2\pi it(k_1 - k_2)}{n} \right)$$

$$= \frac{1}{n} \sum_{r = -(n-1)}^{n-1} c(r) \exp \left( ir \frac{2\pi k_2}{n} \right) \sum_{t=1}^{n} \exp \left( \frac{2\pi it(k_1 - k_2)}{n} \right) \frac{1}{n} \delta_{k_1}(k_2) + R_n,$$

where

$$R_n = \frac{1}{n} \sum_{r = -(n-1)}^{n-1} c(r) \exp \left( -ir \frac{2\pi k_2}{n} \right) \sum_{t=n-|r|+1}^{n} \exp \left( \frac{2\pi it(k_1 - k_2)}{n} \right)$$

240
Thus $|R_n| \leq \frac{1}{n} \sum_{|r| \leq n} |rc(r)| = O(n^{-1})$ Finally by using (8.8) we obtain the result.

Exercise 8.2 The the above proof (in Section 8.2.3) uses that $\sum_r |rc(r)| < \infty$. What bounds do we obtain if we relax this assumption to $\sum_r |c(r)| < \infty$?

8.2.4 Heuristics

In this section we summarize some spectral properties. We do this by considering the DFT of the data $\{J_n(\omega_k)\}_{k=1}^n$. It is worth noting that to calculate $\{J_n(\omega_k)\}_{k=1}^n$ is computationally very fast and requires only $O(n \log n)$ computing operations (see Section A.5, where the Fast Fourier Transform is described).

The spectral (Cramer’s) representation theorem

We observe that for any sequence $\{X_t\}_{t=1}^n$ that it can be written as the inverse transform for $1 \leq t \leq n$

$$X_t = \frac{1}{\sqrt{n}} \sum_{k=1}^n J_n(\omega_k) \exp(-it\omega_k), \quad (8.9)$$

which can be written as an integral

$$X_t = \sum_{k=2}^n \exp(-it\omega_k) [Z_n(\omega_k) - Z_n(\omega_{k-1})] = \int_0^{2\pi} \exp(-it\omega)dZ_n(\omega), \quad (8.10)$$

where $Z_n(\omega) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} J_n(\omega_k)$.

The second order stationary property of $X_t$ means that the DFT $J_n(\omega_k)$ is close to an uncorrelated sequence or equivalently the process $Z_n(\omega)$ has near ‘orthogonal’ increments, meaning that for any two non-intersecting intervals $[\omega_1, \omega_2]$ and $[\omega_3, \omega_4]$ that $Z_n(\omega_2) - Z_n(\omega_1)$ and $Z_n(\omega_4) - Z_n(\omega_3)$. The spectral representation theorem generalizes this result, it states that for any second order stationary time series $\{X_t\}$ there exists an a process $\{Z(\omega); \omega \in [0, 2\pi]\}$ where for all $t \in \mathbb{Z}$

$$X_t = \int_0^{2\pi} \exp(-it\omega)dZ(\omega) \quad (8.11)$$

and $Z(\omega)$ has orthogonal increments, meaning that for any two non-intersecting intervals $[\omega_1, \omega_2]$ and $[\omega_3, \omega_4]$ $\text{E}[Z(\omega_2) - Z(\omega_1)][Z(\omega_2) - Z(\omega_1)] = 0$. 

241
We now explore the relationship between the DFT with the orthogonal increment process. Using (8.11) we see that

\[ J_n(\omega_k) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \exp(it\omega_k) = \frac{1}{\sqrt{2\pi n}} \int_0^{2\pi} \left( \sum_{t=1}^{n} \exp(it[\omega_k - \omega]) \right) dZ(\omega) \]

\[ = \frac{1}{\sqrt{2\pi n}} \int_0^{2\pi} \left( e^{i(n+1)(\omega_k - \omega)/2} D_{n/2}(\omega_k - \omega) \right) dZ(\omega), \]

where \( D_{n/2}(x) = \sin((n + 1)/2)x)/\sin(x/2) \) is the Dirichlet kernel (see Priestley (1983), page 419). We recall that the Dirichlet kernel limits to the Dirac-delta function, therefore very crudely speaking we observe that the DFT is an approximation of the orthogonal increment localized about \( \omega_k \) (though mathematically this is not strictly correct).

**Bochner’s theorem**

This is a closely related result that is stated in terms of the so called spectral distribution. First the heuristics. We see that from Lemma 8.2.1 that the DFT \( J_n(\omega_k) \), is close to uncorrelated. Using this and inverse Fourier transforms we see that for \( 1 \leq t, \tau \leq n \) we have

\[ c(t - \tau) = \text{cov}(X_t, X_\tau) = \frac{1}{n} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \text{cov}(J_n(\omega_{k_1}), J_n(\omega_{k_2})) \exp(-it\omega_{k_1} + i\tau\omega_{k_2}) \]

\[ \approx \frac{1}{n} \sum_{k=1}^{n} \text{var}(J_n(\omega_k)) \exp(-i(t - \tau)\omega_k). \quad (8.12) \]

Let \( F_n(\omega) = \frac{1}{n} \sum_{k=1}^{n} \text{var}(J_n(\omega_k)) \), then the above can be written as

\[ c(t - \tau) \approx \int_0^{2\pi} \exp(-i(t - \tau)\omega)dF_n(\omega), \]

where we observe that \( F_n(\omega) \) is a positive function which in non-decreasing over \( \omega \). Bochner’s theorem is an extension of this is states that for any autocovariance function \( \{c(k)\} \) we have the representation

\[ c(t - \tau) = \int_0^{2\pi} \exp(-i(t - \tau)\omega)f(\omega)d\omega = \int_0^{2\pi} \exp(-i(t - \tau)\omega)dF(\omega), \]

where \( F(\omega) \) is a positive non-decreasing bounded function. Moreover, \( F(\omega) = E(|Z(\omega)|^2) \). We note that if the spectral density function exists (which is only true if \( \sum_r |c(r)|^2 < \infty \) then \( F(\omega) = \)

242
\[ \int_0^\omega f(\lambda)d\lambda. \]

**Remark 8.2.2** The above results hold for both linear and nonlinear time series, however, in the case that \( X_t \) has a linear representation

\[ X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \]

then \( X_t \) has the particular form

\[ X_t = \int A(\omega) \exp(-i\omega) dZ(\omega), \tag{8.13} \]

where \( A(\omega) = \sum_{j=-\infty}^{\infty} \psi_j \exp(ij\omega) \) and \( Z(\omega) \) is an orthogonal increment process, but in addition \( \mathbb{E}(dZ(\omega))^2 = d\omega \) i.e. the variance of increments do not vary over frequency (as this varying has been absorbed by \( A(\omega) \), since \( F(\omega) = |A(\omega)|^2 \)).

We mention that a more detailed discussion on spectral analysis in time series is give in Priestley (1983), Chapters 4 and 6, Brockwell and Davis (1998), Chapters 4 and 10, Fuller (1995), Chapter 3, Shumway and Stoffer (2006), Chapter 4. In many of these references they also discuss tests for periodicity etc (see also Quinn and Hannan (2001) for estimation of frequencies etc.).

8.3 The spectral density and spectral distribution

8.3.1 The spectral density and some of its properties

We start by showing that under certain strong conditions the spectral density function is non-negative. We later weaken these conditions (and this is often called Bochner’s theorem).

**Theorem 8.3.1 (Positiveness of the spectral density)** Suppose the coefficients \( \{c(k)\} \) are absolutely summable (that is \( \sum_k |c(k)| < \infty \)). Then the sequence \( \{c(k)\} \) is positive semi-definite if and only if the function \( f(\omega) \), where

\[ f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} c(k) \exp(ik\omega) \]
is nonnegative. Moreover

\[ c(k) = \int_0^{2\pi} \exp(-ik\omega) f(\omega) d\omega. \tag{8.14} \]

It is worth noting that \( f \) is called the spectral density corresponding to the covariances \( \{c(k)\} \).

PROOF. We first show that if \( \{c(k)\} \) is a non-negative definite sequence, then \( f(\omega) \) is a nonnegative function. We recall that since \( \{c(k)\} \) is non-negative then for any sequence \( x = (x_1, \ldots, x_N) \) (real or complex) we have \( \sum_{s,t=1}^{n} x_s c(s-t) \bar{x}_s \geq 0 \) (where \( \bar{x}_s \) is the complex conjugate of \( x_s \)). Now we consider the above for the particular case \( x = (\exp(i\omega), \ldots, \exp(in\omega)) \). Define the function

\[ f_n(\omega) = \frac{1}{2\pi n} \sum_{s,t=1}^{n} \exp(is\omega) c(s-t) \exp(-it\omega). \]

Thus by definition \( f_n(\omega) \geq 0 \). We note that \( f_n(\omega) \) can be rewritten as

\[ f_n(\omega) = \frac{1}{2\pi} \sum_{k=-n}^{n-1} \left( \frac{n-|k|}{n} \right) c(k) \exp(ik\omega). \]

Comparing \( f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} c(k) \exp(ik\omega) \) with \( f_n(\omega) \) we see that

\[ |f(\omega) - f_n(\omega)| \leq \frac{1}{2\pi} \left| \sum_{|k| \geq n} c(k) \exp(ik\omega) \right| + \frac{1}{2\pi} \left| \sum_{k=-n}^{n-1} \frac{|k|}{n} c(k) \exp(ik\omega) \right| \]

\[ := I_n + II_n. \]

Since \( \sum_{k=-\infty}^{\infty} |c(k)| < \infty \) it is clear that \( I_n \to 0 \) as \( n \to \infty \). Using Lemma A.1.1 we have \( II_n \to 0 \) as \( n \to \infty \). Altogether the above implies

\[ |f(\omega) - f_n(\omega)| \to 0 \quad \text{as} \quad n \to \infty. \tag{8.15} \]

Now it is clear that since for all \( n \), \( f_n(\omega) \) are nonnegative functions, the limit \( f \) must be nonnegative (if we suppose the contrary, then there must exist a sequence of functions \( \{f_{n_k}(\omega)\} \) which are not necessarily nonnegative, which is not true). Therefore we have shown that if \( \{c(k)\} \) is a nonnegative definite sequence, then \( f(\omega) \) is a nonnegative function.

We now show the converse, that is the Fourier coefficients of any non-negative \( \ell_2 \) function \( f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} c(k) \exp(ik\omega) \), is a positive semi-definite sequence. Writing \( c(k) = \int_0^{2\pi} f(\omega) \exp(ik\omega) d\omega \)
we substitute this into Definition 1.6.1 to give

\[ \sum_{s,t=1}^{n} x_s c(s-t)x_t = \int_{0}^{2\pi} f(\omega) \left\{ \sum_{s,t=1}^{n} x_s \exp(i(s-t)\omega)x_t \right\} d\omega = \int_{0}^{2\pi} f(\omega) \left| \sum_{s=1}^{n} x_s \exp(is\omega) \right|^2 d\omega \geq 0. \]

Hence we obtain the desired result. \( \Box \)

The above theorem is very useful. It basically gives a simple way to check whether a sequence \( \{c(k)\} \) is non-negative definite or not (hence whether it is a covariance function - recall Theorem 1.6.1). See Brockwell and Davis (1998), Corollary 4.3.2 or Fuller (1995), Theorem 3.1.9, for alternative explanations.

**Example 8.3.1** Consider the empirical covariances (here we gives an alternative proof to Remark 6.2.1) defined in Chapter 6

\[ \hat{c}_n(k) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-|k|} X_t X_{t+|k|} & |k| \leq n-1 \\ 0 & \text{otherwise} \end{cases} , \]

we give an alternative proof to Lemma 6.2.1 to show that \( \{\hat{c}_n(k)\} \) is non-negative definite sequence. To show that the sequence we take the Fourier transform of \( \hat{c}_n(k) \) and use Theorem 8.3.1. The Fourier transform of \( \{\hat{c}_n(k)\} \) is

\[ \sum_{k=-(n-1)}^{(n-1)} \exp(i\omega)\hat{c}_n(k) = \sum_{k=-(n-1)}^{(n-1)} \exp(i\omega) \frac{1}{n} \sum_{t=1}^{n-|k|} X_t X_{t+|k|} = \frac{1}{n} \sum_{t=1}^{n} X_t \exp(it\omega) \geq 0. \]

Since the above is non-negative, this means that \( \{\hat{c}_n(k)\} \) is a non-negative definite sequence.

We now state a useful result which relates the largest and smallest eigenvalue of the variance of a stationary process to the smallest and largest values of the spectral density (we recall we used this in Lemma 5.4.1).

**Lemma 8.3.1** Suppose that \( \{X_k\} \) is a stationary process with covariance function \( \{c(k)\} \) and spectral density \( f(\omega) \). Let \( \Sigma_n = \text{var}(\bar{X}_n) \), where \( \bar{X}_n = (X_1, \ldots, X_n) \). Suppose \( \inf_{\omega} f(\omega) \geq m > 0 \) and \( \sup_{\omega} f(\omega) \leq M < \infty \). Then for all \( n \) we have

\[ \lambda_{\min}(\Sigma_n) \geq \inf_{\omega} f(\omega) \quad \text{and} \quad \lambda_{\max}(\Sigma_n) \leq \sup_{\omega} f(\omega). \]
PROOF. Let \( \xi_1 \) be the eigenvector with smallest eigenvalue \( \lambda_1 \) corresponding to \( \Sigma_n \). Then using \( c(s-t) = \int f(\omega) \exp(i(s-t)\omega)d\omega \) we have

\[
\lambda_{\min}(\Sigma_n) = \xi_1' \Sigma_n \xi_1 = \sum_{s,t=1}^n \bar{e}_{s,1}c(s-t)e_{t,1} = \int f(\omega) \sum_{s,t=1}^n \bar{e}_{s,1} \exp(i(s-t)\omega)e_{t,1}d\omega = \\
= \int_0^{2\pi} f(\omega) \left| \sum_{s=1}^n e_{s,1} \exp(is\omega) \right|^2 d\omega \geq \inf_{\omega} f(\omega) \int_0^{2\pi} \left| \sum_{s=1}^n e_{s,1} \exp(is\omega) \right|^2 d\omega = \inf_{\omega} f(\omega),
\]

since by definition \( \int \left| \sum_{s=1}^n e_{s,1} \exp(is\omega) \right|^2 d\omega = \sum_{s=1}^n |e_{s,1}|^2 = 1 \) (using Parseval’s identity). Using a similar method we can show that \( \lambda_{\max}(\Sigma_n) \leq \sup f(\omega) \).

We now state a version of the above result which requires weaker conditions on the autocovariance function (only that they decay to zero).

**Lemma 8.3.2** Suppose the covariance \( \{c(k)\} \) decays to zero as \( k \to \infty \), then for all \( n \), \( \Sigma_n = \text{var}(X_n) \) is a non-singular matrix (Note we do not require the stronger condition the covariances are absolutely summable).

PROOF. See Brockwell and Davis (1998), Proposition 5.1.1.

**8.3.2 The spectral distribution and Bochner’s (Hergoltz) theorem**

Theorem 8.3.1 hinges on the result that \( f_n(\omega) = \sum_{r=-(n-1)}^{(n-1)}(1 - |r|/n)e^{ir\omega} \) has a well defined pointwise limit as \( n \to \infty \), this only holds when the sequence \( \{c(k)\} \) is absolutely summable. Of course this may not always be the case. An extreme example is the time series \( X_t = Z \). Clearly this is a stationary time series and its covariance is \( c(k) = \text{var}(Z) = 1 \) for all \( k \). In this case the autocovariance sequence \( \{c(k) = 1; k \in \mathbb{Z}\} \), is not absolutely summable, hence the representation of the covariance in Theorem 8.3.1 does not apply. The reason is because the Fourier transform of the infinite sequence \( \{c(k) = 1; k \in \mathbb{Z}\} \) is not well defined (clearly \( \{c(k) = 1\}_{k} \) does not belong to \( \ell_1 \)).

However, we now show that Theorem 8.3.1 can be generalised to include all non-negative definite sequences and stationary processes, by considering the spectral distribution rather than the spectral density.
Theorem 8.3.2  A function \( \{c(k)\} \) is non-negative definite sequence if and only if

\[
c(k) = \int_0^{2\pi} \exp(-ik\omega)dF(\omega),
\]  

(8.16)

where \( F(\omega) \) is a right-continuous (this means that \( F(x+h) \to F(x) \) as \( 0 < h \to 0 \)), non-decreasing, non-negative, bounded function on \([-\pi, \pi]\) (hence it has all the properties of a distribution and it can be consider as a distribution - it is usually called the spectral distribution). This representation is unique.

This is a very constructive result. It shows that the Fourier coefficients of any distribution function form a non-negative definite sequence, and thus, if \( c(k) = c(-k) \) (hence is symmetric) correspond to the covariance function of a random process. In Figure 8.7 we give two distribution functions. the top plot is continuous and smooth, therefore it’s derivative will exist, be positive and belong to \( \ell_2 \). So it is clear that its Fourier coefficients form a non-negative definite sequence. The interesting aspect of Theorem 8.3.2 is that the Fourier coefficients corresponding to the distribution function in the second plot also forms a non-negative definite sequence even though the derivative of this distribution function does not exist. However, this sequence will not belong to \( \ell_2 \) (ie. the correlations function will not decay to zero as the lag grows).

Figure 8.7: Both plots are of non-decreasing functions, hence are valid distribution functions. The top plot is continuous and smooth, thus its derivative (the spectral density function) exists. Whereas the bottom plot is not (spectral density does not exist).
**PROOF of Theorem 8.3.2.** We first show that if \( \{c(k)\} \) is non-negative definite sequence, then we can write \( c(k) = \int_0^{2\pi} \exp(ik\omega)dF(\omega) \), where \( F(\omega) \) is a distribution function.

To prove the result we adapt some of the ideas used to prove Theorem 8.3.1. As in the proof of Theorem 8.3.1 define the (nonnegative) function

\[
f_n(\omega) = \text{var}[J_n(\omega)] = \frac{1}{2\pi n} \sum_{s,t=1}^{n} \exp(is\omega)c(s-t)\exp(-it\omega) = \frac{1}{2\pi n} \sum_{k=-(n-1)}^{(n-1)} \left( \frac{n-|k|}{n} \right) c(k) \exp(ik\omega).
\]

If \( \{c(k)\} \) is not absolutely summable, the limit of \( f_n(\omega) \) is no longer well defined. Instead we consider its integral, which will always be a distribution function (in the sense that it is nondecreasing and bounded). Let us define the function \( F_n(\omega) \) whose derivative is \( f_n(\omega) \), that is

\[
F_n(\omega) = \int_0^\omega f_n(\lambda)d\lambda = \frac{\omega}{2\pi} c(0) + \frac{1}{2\pi} \sum_{r=1}^{n-1} \left( 1 - \frac{r}{n} \right) c(r) \frac{\sin(\omega r)}{r} \quad 0 \leq \lambda \leq 2\pi.
\]

Since \( f_n(\lambda) \) is nonnegative, \( F_n(\omega) \) is a nondecreasing function. Furthermore it is bounded since

\[
F_n(2\pi) = \int_0^{2\pi} f_n(\lambda)d\lambda = c(0).
\]

Hence \( F_n \) satisfies all properties of a distribution and can be treated as a distribution function. This means that we can apply Helly’s theorem to the sequence \( \{F_n\}_n \). We first recall that if \( \{x_n\} \) are real numbers defined on a compact set \( X \subset \mathbb{R} \), then there exists a subsequence \( \{x_{n_m}\}_m \) which has a limit in the set \( X \) (this is called the Bolzano-Weierstrass theorem). An analogous result exists for measures, this is called Helly’s theorem (see Ash (1972), page 329). It states that for any sequence of distributions \( \{G_n\} \) defined on \([0,2\pi]\), were \( G_n(0) = 0 \) and \( \sup_n G_n(2\pi) < M < \infty \), there exists a subsequence \( \{n_m\}_m \) where \( G_{n_m}(x) \to G(x) \) as \( m \to \infty \) for each \( x \in [0,2\pi] \) at which \( G \) is continuous. Furthermore, since \( G_{n_m}(x) \to G(x) \) (pointwise as \( m \to \infty \)), this implies (see Varadhan, Theorem 4.1 for equivalent forms of convergence) that for any bounded sequence \( h \) we have that

\[
\int h(x)dG_{n_m}(x) \to \int h(x)dG(x) \quad \text{as} \quad m \to \infty.
\]

We now apply this result to \( \{F_n\}_n \). Using Helly’s theorem there exists a subsequence of distributions
\{F_{n_m}\}_m which has a pointwise limit \(F\). Thus for any bounded function \(h\) we have

\[
\int h(x) dF_{n_m}(x) \to \int h(x) dF(x) \quad \text{as } m \to \infty. \tag{8.17}
\]

We focus on the function \(h(x) = \exp(-ik\omega)\). It is clear that for every \(k\) and \(n\) we have

\[
\int_0^{2\pi} \exp(-ik\omega) dF_n(\omega) = \int_0^{2\pi} \exp(ik\omega) f_n(\omega) d\omega = \begin{cases} (1 - \frac{|k|}{n})c(k) & |k| \leq n \\ 0 & |k| \geq n \end{cases} \tag{8.18}
\]

Define the sequence

\[
d_{n,k} = \int_0^{2\pi} \exp(ik\omega) dF_n(\omega) = \left(1 - \frac{|k|}{n}\right)c(k).
\]

We observe that for fixed \(k\), \(\{d_{n,k}; n \in \mathbb{Z}\}\) is a Cauchy sequence, where

\[
d_{n,k} \to d_k = c(k) \tag{8.19}
\]

as \(n \to \infty\).

Now we use (8.17) and focus on the convergent subsequence \(\{n_m\}_m\). By using (8.17) we have

\[
d_{n_m,k} = \int \exp(-ikx) dF_{n_m}(x) \to \int \exp(-ikx) dF(x) \quad \text{as } m \to \infty
\]

and by (8.19) \(d_{n_m,k} \to c(k)\) as \(m \to \infty\). Thus

\[
c(k) = \int \exp(-ikx) dF(x).
\]

This gives the first part of the assertion.

To show the converse, that is \(\{c(k)\}\) is a non-negative definite sequence when \(c(k)\) is defined as \(c(k) = \int \exp(ik\omega)dF(\omega)\), we use the same method given in the proof of Theorem 8.3.1, that is

\[
\sum_{s,t=1}^n x_s c(s-t) \bar{x}_t = \int_0^{2\pi} \left\{ \sum_{s,t=1}^n x_s \exp(-i(s-t)\omega) \bar{x}_t \right\} dF(\omega)
= \int_0^{2\pi} \left| \sum_{s=1}^n x_s \exp(-i\omega) \right|^2 dF(\omega) \geq 0,
\]

since \(F(\omega)\) is a distribution.
Finally, if \( \{c(k)\} \) were absolutely summable, then we can use Theorem 8.3.1 to write \( c(k) = \int_0^{2\pi} \exp(-ik \omega)dF(\omega) \), where \( F(\omega) = \int_0^{\omega} f(\lambda)d\lambda \) and \( f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} c(k) \exp(ik \omega) \). By using Theorem 8.3.1 we know that \( f(\lambda) \) is nonnegative, hence \( F(\omega) \) is a distribution, and we have the result.

\[\Box\]

**Example 8.3.2** Using the above we can construct the spectral distribution for the (rather silly) time series \( X_t = Z \). Let \( F(\omega) = 0 \) for \( \omega < 0 \) and \( F(\omega) = \text{var}(Z) \) for \( \omega \geq 0 \) (hence \( F \) is the step function). Then we have

\[
\text{cov}(X_t, X_{t+k}) = \text{var}(Z) = \int \exp(-ik \omega)dF(\omega).
\]

**Example 8.3.3** Consider the second order stationary time series

\[
X_t = U_1 \cos(\lambda t) + U_2 \sin(\lambda t),
\]

where \( U_1 \) and \( U_2 \) are iid random variables with mean zero and variance \( \sigma^2 \) and \( \lambda \) the frequency. It can be shown that

\[
\text{cov}(X_t, X_{t+k}) = \frac{\sigma^2}{2} \left[ \exp(i\lambda k) + \exp(-i\lambda k) \right].
\]

Observe that this covariance does not decay with the lag \( k \). Then

\[
\text{cov}(X_t, X_{t+k}) = \text{var}(Z) = \int_0^{2\pi} \exp(-ik \omega)dF(\omega).
\]

where

\[
F(\omega) = \begin{cases} 
0 & \omega < -\lambda \\
\sigma^2/2 & -\lambda \leq \omega < \lambda \\
\sigma^2 & \lambda \geq \omega.
\end{cases}
\]

### 8.4 The spectral representation theorem

We now state the spectral representation theorem and give a rough outline of the proof.

**Theorem 8.4.1** If \( \{X_t\} \) is a second order stationary time series with mean zero, and spectral distribution \( F(\omega) \), and the spectral distribution function is \( F(\omega) \), then there exists a right continuous,
orthogonal increment process \{Z(\omega)\} (that is \(E[(Z(\omega_1) - Z(\omega_2))(Z(\omega_3) - Z(\omega_4))] = 0\), when the intervals \([\omega_1, \omega_2]\) and \([\omega_3, \omega_4]\) do not overlap) such that

\[
X_t = \int_0^{2\pi} \exp(-it\omega) dZ(\omega),
\]

where for \(\omega_1 \geq \omega_2\), \(E|Z(\omega_1) - Z(\omega_2)|^2 = F(\omega_1) - F(\omega_2)\) (noting that \(F(0) = 0\)). (One example of a right continuous, orthogonal increment process is Brownian motion, though this is just one example, and usually \(Z(\omega)\) will be far more general than Brownian motion).

Heuristically we see that (8.20) is the decomposition of \(X_t\) in terms of frequencies, whose amplitudes are orthogonal. In other words \(X_t\) is decomposed in terms of frequencies \(\exp(it\omega)\) which have the orthogonal amplitudes \(dZ(\omega) \approx (Z(\omega + \delta) - Z(\omega))\).

Remark 8.4.1 Note that so far we have not defined the integral on the right hand side of (8.20). It is known as a stochastic integral. Unlike many deterministic functions (functions whose derivative exists), one cannot really suppose \(dZ(\omega) \approx Z'(\omega)d\omega\), because usually a typical realisation of \(Z(\omega)\) will not be smooth enough to differentiate. For example, it is well known that Brownian is quite ‘rough’, that is a typical realisation of Brownian motion satisfies \(|B(t_1, \tilde{\omega}) - B(t_2, \tilde{\omega})| \leq K(\tilde{\omega})|t_1 - t_2|\gamma\), where \(\tilde{\omega}\) is a realisation and \(\gamma \leq 1/2\), but in general \(\gamma\) will not be larger. The integral \(\int g(\omega)dZ(\omega)\) is well defined if it is defined as the limit (in the mean squared sense) of discrete sums. More precisely, let \(Z_n(\omega) = \sum_{k=1}^{n} Z(\omega_k)I_{\omega_{k-1}\omega_{k}}(\omega) = \sum_{k=1}^{[\omega/2\pi]} [Z(\omega_k) - Z(\omega_{k-1})]\), then

\[
\int g(\omega)dZ_n(\omega) = \sum_{k=1}^{n} g(\omega_k)\{Z(\omega_k) - Z(\omega_{k-1})\}.
\]

The limit of \(\int g(\omega)dZ_n(\omega)\) as \(n \to \infty\) is \(\int g(\omega)dZ(\omega)\) (in the mean squared sense, that is \(E[\int g(\omega)dZ(\omega) - \int g(\omega)dZ_n(\omega)]^2\)). Compare this with our heuristics in equation (8.10).

For a more precise explanation, see Parzen (1959), Priestley (1983), Sections 3.6.3 and Section 4.11, page 254, and Brockwell and Davis (1998), Section 4.7. For a very good review of elementary stochastic calculus see Mikosch (1999).

A very elegant explanation on the different proofs of the spectral representation theorem is given in Priestley (1983), Section 4.11. We now give a rough outline of the proof using the functional theory approach.
Rough PROOF of the Spectral Representation Theorem To prove the result we first define two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, where $\mathcal{H}_1$ one contains deterministic functions and $\mathcal{H}_2$ contains random variables.

First we define the space

$$\mathcal{H}_1 = \overline{\text{sp}\{e^{it\omega}; t \in \mathbb{Z}\}}$$

with inner-product

$$\langle f, g \rangle = \int_0^{2\pi} f(x)\overline{g(x)}dF(x) \quad (8.21)$$

(and of course distance $\langle f - g, f - g \rangle = \int_0^{2\pi} |f(x) - g(x)|^2dF(x)$) it is clear that this inner product is well defined because $\langle f, f \rangle \geq 0$ (since $F$ is a measure). It can be shown (see Brockwell and Davis (1998), page 144) that $\mathcal{H}_1 = \left\{ g; \int_0^{2\pi} |g(\omega)|^2dF(\omega) < \infty \right\}$. We also define the space

$$\mathcal{H}_2 = \overline{\text{sp}\{X_t; t \in \mathbb{Z}\}}$$


Now let us define the linear mapping $T : \mathcal{H}_1 \to \mathcal{H}_2$

$$T(\sum_{j=1}^{n} a_j \exp(i\omega)) = \sum_{j=1}^{n} a_j X_k, \quad (8.22)$$

for any $n$ (it is necessary to show that this can be extended to infinite $n$, but we won’t do so here). We will shown that $T$ defines an isomorphism (ie. it is a one-to-one linear mapping that preserves norm). To show that it is a one-to-one mapping see Brockwell and Davis (1998), Section 4.7. It is clear that it is linear, there all that remains is to show that the mapping preserves inner-product. Suppose $f, g \in \mathcal{H}_1$, then there exists coefficients $\{f_j\}$ and $\{g_j\}$ such that $f(x) = \sum_j f_j \exp(ij\omega)$ and $g(x) = \sum_j g_j \exp(ij\omega)$. Hence by definition of $T$ in (8.22) we have

$$\langle Tf, Tg \rangle = \text{cov}(\sum_j f_j X_j, \sum_j g_j X_j) = \sum_{j_1, j_2} f_{j_1} \overline{g_{j_2}} \text{cov}(X_{j_1}, X_{j_2}) \quad (8.23)$$

Roughly speaking it is because all continuous functions on $[0, 2\pi]$ are dense in $L_2([0, 2\pi], B, F)$ (using the metric $||f - g|| = \langle f - g, f - g \rangle$ and the limit of Cauchy sequences). Since all continuous function can be written as linear combinations of the Fourier basis, this gives the result.
Now by using Bochner’s theorem (see Theorem 8.3.2) we have

\[
\langle Tf, Tg \rangle = \int_0^{2\pi} \left( \sum_{j_1,j_2} f_{j_1} g_{j_2} \exp(i(j_1 - j_2)\omega) \right) dF(\omega) = \int_0^{2\pi} f(x) \overline{g(x)} dF(x) = \langle f, g \rangle.
\]

(8.24)

Hence \( \langle Tf, Tg \rangle = \langle f, g \rangle \), so the inner product is preserved (hence \( T \) is an isometry).

Altogether this means that \( T \) defines an isomorphism between \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \). Therefore all functions which are in \( \mathcal{H}_1 \) have a corresponding random variable in \( \mathcal{H}_2 \) which has similar properties.

For all \( \omega \in [0, 2\pi] \), it is clear that the identity functions \( I_{[0,\omega]}(x) \in \mathcal{H}_1 \). Thus we define the random function \( \{Z(\omega); 0 \leq \omega \leq 2\pi\} \), where \( T(I_{[0,\omega]}(\cdot)) = Z(\omega) \in \mathcal{H}_2 \) (since \( T \) is an isomorphism).

Since that mapping \( T \) is linear we observe that

\[
T(I_{[\omega_1,\omega_2]}) = T(I_{[0,\omega_1]} - I_{[0,\omega_2]}) = T(I_{[0,\omega_1]}) - T(I_{[0,\omega_2]}) = Z(\omega_1) - Z(\omega_2).
\]

Moreover, since \( T \) preserves the norm for any non-intersecting intervals \( [\omega_1, \omega_2] \) and \( [\omega_3, \omega_4] \) we have

\[
\text{cov} \left( (Z(\omega_1) - Z(\omega_2), (Z(\omega_3) - Z(\omega_4) = \langle T(I_{[\omega_1,\omega_2]}), T(I_{[\omega_3,\omega_4]} \rangle = \langle I_{[\omega_1,\omega_2]}, I_{[\omega_3,\omega_4]} \rangle
\]

\[
= \int I_{[\omega_1,\omega_2]}(\omega) I_{[\omega_3,\omega_4]}(\omega) dF(\omega) = 0.
\]

Therefore by construction \( \{Z(\omega); 0 \leq \omega \leq 2\pi\} \) is an orthogonal increment process, where

\[
\mathbb{E} |Z(\omega_2) - Z(\omega_1)|^2 = \langle T(I_{[\omega_1,\omega_2]}), T(I_{[\omega_1,\omega_2]} \rangle = \langle I_{[\omega_1,\omega_2]}, I_{[\omega_1,\omega_2]} \rangle
\]

\[
= \int_0^{2\pi} I_{[\omega_1,\omega_2]}(\omega) dF(\omega) = \int_{\omega_1}^{\omega_2} dF(\omega) = F(\omega_2) - F(\omega_1).
\]

Having defined the two spaces which are isomorphic and the random function \( \{Z(\omega); 0 \leq \omega \leq 2\pi\} \) and function \( I_{[0,\omega]}(x) \) which have orthogonal increments, we can now prove the result. Since \( dI_{[0,\omega]}(s) = \delta_\omega(s) ds \), where \( \delta_\omega(s) \) is the dirac delta function, any function \( g \in L_2[0, 2\pi] \) can be represented as

\[
g(\omega) = \int_0^{2\pi} g(s) dI_{[\omega,2\pi]}(s).
\]

253
Thus for $g(\omega) = \exp(-it\omega)$ we have

$$\exp(-it\omega) = \int_0^{2\pi} \exp(-its) dI_{[\omega, 2\pi]}(s).$$

Therefore

$$T(\exp(-it\omega)) = T\left(\int_0^{2\pi} \exp(-its) dI_{[\omega, 2\pi]}(s)\right) = \int_0^{2\pi} \exp(-its) T[dI_{[\omega, 2\pi]}(s)]$$

$$= \int_0^{2\pi} \exp(-its) dT[I_{[\omega, 2\pi]}(s)],$$

where the mapping goes inside the integral due to the linearity of the isomorphism. Using that $I_{[\omega, 2\pi]}(s) = I_{[0, s]}(\omega)$ we have

$$T(\exp(-it\omega)) = \int_0^{2\pi} \exp(-its) dT[I_{[0, s]}(\omega)].$$

By definition we have $T(I_{[0, s]}(\omega)) = Z(s)$ which we substitute into the above to give

$$X_t = \int_0^{2\pi} \exp(-its) dZ(s),$$

which gives the required result.

Note that there are several different ways to prove this result. □

It is worth taking a step back from the proof and see where the assumption of stationarity crept in. By Bochner’s theorem we have that

$$c(t - \tau) = \int \exp(-i(t - \tau)\omega) dF(\omega),$$

where $F$ is a distribution. We use $F$ to define the space $\mathcal{H}_1$, the mapping $T$ (through $\{\exp(ik\omega)\}_k$), the inner-product and thus the isomorphism. However, it was the construction of the orthogonal random functions $\{Z(\omega)\}$ that was instrumental. The main idea of the proof was that there are functions $\{\phi_k(\omega)\}$ and a distribution $H$ such that all the covariances of the stochastic process $\{X_t\}$ can be written as

$$E(X_t X_\tau) = c(t, \tau) = \int_0^{2\pi} \phi_t(\omega) \overline{\phi_\tau(\omega)} dH(\omega),$$

254
where $H$ is a measure. As long as the above representation exists, then we can define two spaces $H_1$ and $H_2$ where $\{\phi_k\}$ is the basis of the functional space $H_1$ and it contains all functions $f$ such that $\int |f(\omega)|^2 dH(\omega) < \infty$ and $H_2$ is the random space defined by $\mathbb{P}(X_t; t \in \mathbb{Z})$. From here we can define an isomorphism $T : H_1 \rightarrow H_2$, where for all functions $f(\omega) = \sum k f_k \phi_k(\omega) \in H_1$

$$T(f) = \sum_k f_k X_k \in H_2.$$ 

An important example is $T(\phi_k) = X_k$. Now by using the same arguments as those in the proof above we have

$$X_t = \int \phi_t(\omega) dZ(\omega)$$

where $\{Z(\omega)\}$ are orthogonal random functions and $E|Z(\omega)|^2 = H(\omega)$. We state this result in the theorem below (see Priestley (1983), Section 4.11).

**Theorem 8.4.2 (General orthogonal expansions)** Let $\{X_t\}$ be a time series (not necessarily second order stationary) with covariance $\{E(X_t X_r) = c(t, s)\}$. If there exists a sequence of functions $\{\phi_k(\cdot)\}$ which satisfy for all $k$

$$\int_0^{2\pi} |\phi_k(\omega)|^2 dH(\omega) < \infty$$

and the covariance admits the representation

$$c(t, s) = \int_0^{2\pi} \phi_t(\omega) \overline{\phi_s(\omega)} dH(\omega), \quad (8.25)$$

where $H$ is a distribution then for all $t$ we have the representation

$$X_t = \int \phi_t(\omega) dZ(\omega) \quad (8.26)$$

where $\{Z(\omega)\}$ are orthogonal random functions and $E|Z(\omega)|^2 = H(\omega)$. On the other hand if $X_t$ has the representation $(8.26)$, then $c(s, t)$ admits the representation $(8.25)$.

**Remark 8.4.2** We mention that the above representation applies to both stationary and nonstationary time series. What makes the exponential functions $\{\exp(ik\omega)\}$ special is if a process is
stationary then the representation of $c(k) := \text{cov}(X_t, X_{t+k})$ in terms of exponentials is guaranteed:

$$c(k) = \int_{0}^{2\pi} \exp(-ik\omega)dF(\omega). \quad (8.27)$$

Therefore there always exists an orthogonal random function $\{Z(\omega)\}$ such that

$$X_t = \int \exp(-it\omega)dZ(\omega).$$

Indeed, whenever the exponential basis is used in the definition of either the covariance or the process $\{X_t\}$, the resulting process will always be second order stationary.

Brockwell and Davis (1998), Proposition 4.8.2 states an interesting consequence of the spectral representation theorem. Suppose that $\{X_t\}$ is a second order stationary time series with spectral distribution $F(\omega)$. If $F(\omega)$ has a discontinuity at $\lambda_0$, then $X_t$ almost surely has the representation

$$X_t = \int_{0}^{2\pi} e^{it\omega}dZ(\omega) + e^{it\lambda_0} \left( Z(\lambda_0^+) - Z(\lambda_0^-) \right)$$

where $Z(\lambda_0^-)$ and $Z(\lambda_0^+)$ denote the left and right limit. This result means that discontinuities in the spectral distribution mean that the corresponding time series contains a deterministic sinusoid functions i.e.

$$X_t = A \cos(\lambda_0 t) + B \sin(\lambda_0 t) + \varepsilon_t$$

where $\varepsilon_t$ is a stationary time series. We came across this “feature” in Section 1.2.4. If the spectral distribution contains a discontinuity, then “formally” the spectral density (which is the derivative of the spectral distribution) is the dirac-delta function at the discontinuity. The periodogram is a “crude” (inconsistent) estimator of the spectral density function, however it captures the general features of the underlying spectral density. Look at Figures 1.8-1.10, observe that there is a large peak corresponding the deterministic frequency and that this peak grows taller as the sample size $n$ grows. This large peak is limiting to the dirac delta function.

Finally we state Brockwell and Davis (1998), Proposition 4.9.1, which justifies our use of the DFT. Brockwell and Davis (1998), Proposition 4.9.1 states that if $\{X_t\}$ is a second order stationary
time series with spectral distribution $F$ and $\nu_1$ and $\nu_2$ are continuity points of $F$ then

$$\frac{1}{2\pi} \sum_{|t| \leq n} X_t \int_{\nu_1}^{\nu_2} \exp(it\omega)d\omega \to Z(\nu_2) - Z(\nu_1),$$

where the convergence is in mean squared.

Let $\omega_k = 2\pi k/n$, then using this result we have

$$\frac{1}{2\pi \sqrt{n}} \sum_{|t| \leq n} X_t \exp(it\omega_k) \approx \sqrt{n} \sum_{|t| \leq n} X_t \int_{\omega_k}^{\omega_{k+1}} \exp(it\omega)d\omega \approx \sqrt{n} [Z(\omega_{k+1}) - Z(\omega_k)],$$

without the scaling factor $\sqrt{n}$, the above would limit to zero. Thus as claimed previously, the DFT estimates the “increments”.

### 8.5 The spectral density functions of MA, AR and ARMA models

We obtain the spectral density function for MA($\infty$) processes. Using this we can easily obtain the spectral density for ARMA processes. Let us suppose that $\{X_t\}$ satisfies the representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \tag{8.28}$$

where $\{\varepsilon_t\}$ are iid random variables with mean zero and variance $\sigma^2$ and $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. We recall that the covariance of above is

$$c(k) = \text{E}(X_tX_{t+k}) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k}. \tag{8.29}$$

Since $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, it can be seen that

$$\sum_k |c(k)| \leq \sum_k \sum_{j=-\infty}^{\infty} |\psi_j| \cdot |\psi_{j+k}| < \infty.$$

Hence by using Theorem 8.3.1, the spectral density function of $\{X_t\}$ is well defined. There are several ways to derive the spectral density of $\{X_t\}$, we can either use (8.29) and $f(\omega) = \frac{1}{2\pi} \sum_k c(k) \exp(ik\omega)$ or obtain the spectral representation of $\{X_t\}$ and derive $f(\omega)$ from the spec-
8.5.1 The spectral representation of linear processes

Since \( \{ \varepsilon_t \} \) are iid random variables, using Theorem 8.4.1 there exists an orthogonal random function \( \{ Z(\omega) \} \) such that

\[
\varepsilon_t = \int_0^{2\pi} \exp(-it\omega) dZ(\omega).
\]

Since \( E(\varepsilon_t) = 0 \) and \( E(\varepsilon_t^2) = \sigma^2 \) multiplying the above by \( \varepsilon_t \), taking expectations and noting that due to the orthogonality of \( \{ Z(\omega) \} \) we have \( E(dZ(\omega_1)d\overline{Z(\omega_2)}) = 0 \) unless \( \omega_1 = \omega_2 \) we have that \( E(|dZ(\omega)|^2) = \sigma^2 d\omega \), hence \( f_\varepsilon(\omega) = (2\pi)^{-1}\sigma^2 \).

Using the above we obtain the following spectral representation for \( \{ X_t \} \)

\[
X_t = \int_0^{2\pi} \left\{ \sum_{j=-\infty}^{\infty} \psi_j \exp(ij\omega) \right\} \exp(-it\omega) dZ(\omega).
\]

Hence

\[
X_t = \int_0^{2\pi} A(\omega) \exp(-it\omega) dZ(\omega),
\]

where \( A(\omega) = \sum_{j=-\infty}^{\infty} \psi_j \exp(ij\omega) \), noting that this is the unique spectral representation of \( X_t \).

**Definition 8.5.1 (The Cramer Representation)** We mention that the representation in (8.30) of a stationary process is usually called the Cramer representation of a stationary process, where

\[
X_t = \int_0^{2\pi} A(\omega) \exp(-it\omega) dZ(\omega),
\]

where \( \{ Z(\omega) : 0 \leq \omega \leq 2\pi \} \) are orthogonal functions.

**Exercise 8.3**

(i) Suppose that \( \{ X_t \} \) has an MA(1) representation \( X_t = \theta \varepsilon_t + \varepsilon_{t-1} \). What is its Cramer’s representation?

(ii) Suppose that \( \{ X_t \} \) has a causal AR(1) representation \( X_t = \phi X_{t-1} + \varepsilon_t \). What is its Cramer’s representation?
8.5.2 The spectral density of a linear process

Multiplying (8.30) by $X_{t+k}$ and taking expectations gives

$$\mathbb{E}(X_t X_{t+k}) = c(k) = \int_0^{2\pi} A(\omega_1)A(-\omega_2)\exp(-i(t+k)\omega_1 + it\omega_2)\mathbb{E}(dZ(\omega_1)d\overline{Z(\omega_2)}).$$

Due to the orthogonality of $\{Z(\omega)\}$ we have $\mathbb{E}(dZ(\omega_1)d\overline{Z(\omega_2)}) = 0$ unless $\omega_1 = \omega_2$, altogether this gives

$$\mathbb{E}(X_t X_{t+k}) = c(k) = \int_0^{2\pi} |A(\omega)|^2 \exp(-ik\omega)\mathbb{E}(|dZ(\omega)|^2) = \int_0^{2\pi} f(\omega) \exp(-ik\omega)d\omega,$$

where $f(\omega) = \frac{\sigma^2}{2\pi} |A(\omega)|^2$. Comparing the above with (8.14) we see that $f(\cdot)$ is the spectral density function.

The spectral density function corresponding to the linear process defined in (8.28) is

$$f(\omega) = \frac{\sigma^2}{2\pi} |\sum_{j=-\infty}^{\infty} \psi_j \exp(-ij\omega)|^2.$$

Remark 8.5.1 (An alternative, more hands on proof) An alternative proof which avoids the Cramer representation is to use that the acf of a linear time series is $c(r) = \sigma^2 \sum_k \psi_j \psi_{j+r}$ (see Lemma 3.1.1). Thus by definition the spectral density function is

$$f(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c(r) \exp(ir\omega) = \frac{\sigma^2}{2\pi} \sum_{r=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+r} \exp(i\omega r).$$

Now make a change of variables $s = j + r$ this gives

$$f(\omega) = \frac{\sigma^2}{2\pi} \sum_{s=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_s \exp(i(s-j)\omega) = \frac{\sigma^2}{2\pi} \left| \sum_{j=-\infty}^{\infty} \psi_j e^{ij\omega} \right|^2 = \frac{\sigma^2}{2\pi} |A(\omega)|^2.$$

Example 8.5.1 Let us suppose that $\{X_t\}$ is a stationary ARMA($p,q$) time series (not necessarily invertible or causal), where

$$X_t - \sum_{j=1}^{p} \psi_j X_{t-j} = \sum_{j=1}^{q} \theta_j \varepsilon_{t-j},$$
\{\varepsilon_t\} are iid random variables with \(E(\varepsilon_t) = 0\) and \(E(\varepsilon_t^2) = \sigma^2\). Then the spectral density of \(\{X_t\}\) is

\[
f(\omega) = \frac{\sigma^2}{2\pi} \frac{|1 + \sum_{j=1}^q \theta_j \exp(i j \omega)|^2}{|1 - \sum_{j=1}^q \phi_j \exp(i j \omega)|^2}
\]

We note that because the ARMA is the ratio of trigonometric polynomials, this is known as a rational spectral density.

**Remark 8.5.2** The roots of the characteristic function of an AR process will have an influence on the location of peaks in its corresponding spectral density function. To see why consider the AR(2) model

\[X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t,\]

where \(\{\varepsilon_t\}\) are iid random variables with zero mean and \(E(\varepsilon_t^2) = \sigma^2\). Suppose the roots of the characteristic polynomial \(\phi(B) = 1 - \phi_1 B - \phi_2 B^2\) lie outside the unit circle and are complex conjugates where \(\lambda_1 = r \exp(i \theta)\) and \(\lambda_2 = r \exp(-i \theta)\). Then the spectral density function is

\[
f(\omega) = \frac{\sigma^2}{\pi} \frac{|1 - r \exp(i(\theta - \omega))|^2}{|1 - r \exp(-i(\theta - \omega))|^2} \frac{|1 + r^2 - 2r \cos(\theta - \omega)||1 + r^2 - 2r \cos(\theta + \omega)|}{|1 + r^2 - 2r \cos(\theta - \omega)||1 + r^2 - 2r \cos(\theta + \omega)|}.
\]

If \(r > 0\), the \(f(\omega)\) is maximum when \(\omega = \theta\), on the other hand if, \(r < 0\) then the above is maximum when \(\omega = \theta - \pi\). Thus the peaks in \(f(\omega)\) correspond to peaks in the pseudo periodicities of the time series and covariance structure (which one would expect), see Section 3.1.2. How pronounced these peaks are depend on how close \(r\) is to one. The close \(r\) is to one the larger the peak. We can generalise the above argument to higher order Autoregressive models, in this case there may be multiple peaks. In fact, this suggests that the larger the number of peaks, the higher the order of the AR model that should be fitted.
8.5.3 Approximations of the spectral density to AR and MA spectral densities

In this section we show that the spectral density

\[ f(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c(r) \exp(ir\omega) \]

can be approximated to any order by the spectral density of an AR\((p)\) or MA\((q)\) process.

We do this by truncating the infinite number of covariances by a finite number, however, this does not necessarily lead to a positive definite spectral density. This can easily be proven by noting that

\[ \tilde{f}_m(\omega) = \sum_{r=-m}^{m} c(r) \exp(ir\omega) = \int_{0}^{2\pi} f(\lambda)D_m(\omega - \lambda)d\lambda, \]

where \(D_m(\lambda) = \sin[(n + 1/2)\lambda]/\sin(\lambda/2)\). Observe that \(D_m(\cdot)\) can be negative, which means that \(\tilde{f}_m(\omega)\) can be negative despite \(f\) being positive.

**Example 8.5.2** Consider the AR\((1)\) process \(X_t = 0.75X_{t-1} + \varepsilon_t\) where \(\text{var}[\varepsilon_t] = 1\). In Lemma 3.1.1 we showed that the autocovariance corresponding to this model is \(c(r) = [1 - 0.75^2]^{-1}0.75^{|r|}\).

Let us define a process whose autocorrelation is \(\hat{c}(0) = [1 - 0.75^2]^{-1}, \hat{c}(1) = c(-1) = [1 - 0.75^2]^{-1}0.75\) and \(\hat{c}(r) = 0\) for \(|r| > 1\). The ‘spectral density’ of this process is

\[ \tilde{f}_m(\omega) = \frac{1}{1 - 0.75^2} \left(1 + 2 \times \frac{3}{4} \cos[\omega]\right). \]

It is clear that this function can be zero for some values of \(\omega\). This means that \(\{\hat{c}(r)\}\) is not a well defined covariance function, hence there does not exist a time series with this covariance structure. In other words, simply truncating an autocovariance is not enough to guarantee that it positive definite sequence.

Instead we consider a slight variant on this and define

\[ \frac{1}{2\pi} \sum_{r=-m}^{m} \left(1 - \frac{|r|}{m}\right) c(r) \exp(ir\omega) \]

which is positive.
Remark 8.5.3 We note that $f_m$ is known as a Cesáro sum because it can be written as

$$f_m(\omega) = \frac{1}{2\pi} \sum_{r=-m}^{m} \left(1 - \frac{|r|}{m}\right) c(r) \exp(ir\omega) = \frac{1}{m} \sum_{n=0}^{m} \tilde{f}_n(\omega), \quad (8.31)$$

where $\tilde{f}_n(\cdot) = \frac{1}{2\pi} \sum_{r=-n}^{n} c(r) \exp(ir\omega)$. Strangely, there is no guarantee that the truncated Fourier transform $\tilde{f}_n$ is not negative, however $f_n(\cdot)$ is definitely positive. There are are a few ways to prove this:

(i) The first method we came across previously, $\text{var}[J_n(\omega)] = f_n(\omega)$, it is clear that using this construction $\inf_{\omega} f_n(\omega) \geq 0$.

(ii) By using (8.31) we can write $f_m(\cdot)$ as

$$f_m(\omega) = \int_0^{2\pi} f(\lambda) F_m(\omega - \lambda) d\lambda,$$

where $F_m(\lambda) = \frac{1}{m} \sum_{r=-m}^{m} D_r(\lambda) = \frac{1}{m} \left(\frac{\sin(n\lambda/2)}{\sin(\lambda/2)}\right)^2$ and $D_r(\lambda) = \sum_{j=-r}^{r} \exp(ij\omega)$ (these are the Fejer and Dirichlet kernels respectively). Since both $f$ and $F_m$ are positive, then $f_m$ has to be positive.

The Cesaro sum is special in the sense that

$$\sup_{\omega} |f_m(\omega) - f(\omega)| \to 0, \quad \text{as } m \to \infty. \quad (8.32)$$

Thus for a large enough $m$, $f_m(\omega)$ will be within $\delta$ of the spectral density $f$. Using this we can prove the results below.

Lemma 8.5.1 Suppose that $\sum_r |c(r)| < \infty$, $f$ is the spectral density of the covariances and $\inf_{\omega \in [0,2\pi]} f(\omega) > 0$. Then for every $\delta > 0$, there exists a $m$ such that $|f(\omega) - f_m(\omega)| < \delta$ and $f_m(\omega) = \sigma^2 |\psi(\omega)|^2$, where $\psi(\omega) = \sum_{j=0}^{m} \psi_j \exp(ij\omega)$. Thus we can approximate the spectral density of $f$ with the spectral density of a MA.

PROOF. We show that there exists an MA($m$) which has the spectral density $f_m(\omega)$, where $f_m$ is defined in (8.31). Thus by (8.32) we have the result.

Before proving the result we note that if a “polynomial” is of the form

$$p(z) = a_0 + \sum_{j=1}^{m} a_j (z + z^{-1})$$

262
then it has the factorization $p(z) = C \prod_{j=1}^{m}[1 - \lambda_j z][1 - \lambda_j^{-1}z]$, where $\lambda_j$ is such that $|\lambda_j| < 1$. Furthermore, if $\{a_j\}_{j=0}^{m}$ are real and $z^m p(z)$ has no roots on the unit circle, then the coefficients of the polynomial $\prod_{j=1}^{m}[1 - \lambda_j z]$ are real. The above claims are true because

(i) To prove that $p(z) = C \prod_{j=1}^{m}[1 - \lambda_j z][1 - \lambda_j^{-1}z]$, we note that $z^m p(z)$ is a $2m$-order polynomial. Thus it can be factorized. If there exists a root $\lambda$ whose inverse is not a root, then the resulting polynomial will have not have the symmetric structure.

(ii) By the complex conjugate theorem, since $z^m p(z)$ has real coefficients, then its complex roots must be conjugates. Moreover, since no roots lie on the unit circle, then no conjugates lie on the unit circle. Thus the coefficients of $\prod_{j=1}^{m}[1 - \lambda_j z]$ are real (if it did lie on the unit circle, then we can distribute the two roots between the two polynomials).

Thus setting $z = e^{i\omega}$

$$\sum_{r=-m}^{m} a_r \exp(i r \omega) = C \prod_{j=1}^{m} [1 - \lambda_j \exp(i \omega)][1 - \lambda_j^{-1} \exp(-i \omega)],$$

for some finite constant $C$. We use the above result. Since $\inf f_m(\omega) > 0$ and setting $a_r = [1 - |r| n^{-1}] c(r)$, we can write $f_m$ as

$$f_m(\omega) = K \left[ \prod_{j=1}^{m} (1 - \lambda_j^{-1} \exp(i \omega)) \right] \left[ \prod_{j=1}^{m} (1 - \lambda_j \exp(-i \omega)) \right]$$

$$= A(\omega) A(-\omega) = |A(\omega)|^2,$$

where

$$A(z) = \prod_{j=1}^{m} (1 - \lambda_j^{-1}z).$$

Since $A(z)$ is an $m$th order polynomial where all the roots are greater than 1, we can always construct an MA($m$) process which has $A(z)$ as its ‘transfer’ function. Thus there exists an MA($m$) process which has $f_m(\omega)$ as its spectral density function.

Remark 8.5.4 (i) The above result requires that $\inf \omega f(\omega) > 0$, in order to ensure that $f_m(\omega)$ is strictly positive. This assumption can be relaxed (and the proof becomes a little more complicated), see Brockwell and Davis (1998), Theorem 4.4.3.

263
Lemma 8.5.2 Suppose that \( \sum_r |c(r)| < \infty \) and \( f \) is corresponding the spectral density function where \( \inf_\omega f(\omega) > 0 \). Then for every \( \delta > 0 \), there exists a \( m \) such that \( |f(\omega) - g_m(\omega)| < \delta \) and \( g_m(\omega) = \sigma^2 |\phi(\omega)^{-1}|^2 \), where \( \phi(\omega) = \sum_{j=0}^{m} \phi_j \exp(\bar{i}\omega) \) and the roots of \( \phi(z) \) lie outside the unit circle. Thus we can approximate the spectral density of \( f \) with the spectral density of a causal autoregressive process.

PROOF. We first note that we can write

\[
|f(\omega) - g_m(\omega)| = f(\omega)|g_m(\omega)^{-1} - f(\omega)^{-1}|g_m(\omega).
\]

Since \( f(\cdot) \in L_2 \) and is bounded away from zero, then \( f^{-1} \in L_2 \) and we can write \( f^{-1} \) as

\[
f^{-1}(\omega) = \sum_{r=\infty}^{\infty} d_r \exp(\bar{i}\omega),
\]

where \( d_r \) are the Fourier coefficients of \( f^{-1} \). Since \( f \) is positive and symmetric, then \( f^{-1} \) is positive and symmetric such that \( f^{-1}(\omega) = \sum_{r=-\infty}^{\infty} d_r \exp(\bar{i}\omega) \) and \( \{d_r\} \) is a positive definite symmetric sequence. Thus we can define the positive function \( g_m \) where

\[
g_m^{-1}(\omega) = \sum_{|r| \leq m} \left( 1 - \frac{|r|}{m} \right) d_r \exp(\bar{i}\omega)
\]

and is such that \( |g_m^{-1}(\omega) - f^{-1}(\omega)| < \delta \), which implies

\[
|f(\omega) - g_m(\omega)| \leq \left( \sum_r |c(r)| \right)^2 \delta.
\]

Now we can apply the same arguments to prove to Lemma 8.5.1 we can show that \( g_m^{-1} \) can be factorised as \( g_m^{-1}(\omega) = C|\phi_m(\omega)|^2 \) (where \( \phi_m \) is an \( m \)th order polynomial whose roots lie outside the unit circle). Thus \( g_m(\omega) = C|\phi_m(\omega)|^{-2} \) and we obtain the desired result.

\[\square\]

8.6 Higher order spectrums

We recall that the covariance is a measure of linear dependence between two random variables. Higher order cumulants are a measure of higher order dependence. For example, the third order
The cumulant for the zero mean random variables $X_1, X_2, X_3$ is

$$\text{cum}(X_1, X_2, X_3) = E(X_1X_2X_3)$$

and the fourth order cumulant for the zero mean random variables $X_1, X_2, X_3, X_4$ is

$$\text{cum}(X_1, X_2, X_3, X_4) = E(X_1X_2X_3X_4) - E(X_1X_2)E(X_3X_4) - E(X_1X_3)E(X_2X_4) - E(X_1X_4)E(X_2X_3).$$

From the definition we see that if $X_1, X_2, X_3, X_4$ are independent then $\text{cum}(X_1, X_2, X_3) = 0$ and $\text{cum}(X_1, X_2, X_3, X_4) = 0$. Moreover, if $X_1, X_2, X_3, X_4$ are Gaussian random variables then $\text{cum}(X_1, X_2, X_3) = 0$ and $\text{cum}(X_1, X_2, X_3, X_4) = 0$. Indeed all cumulants higher than order two is zero. This comes from the fact that cumulants are the coefficients of the power series expansion of the logarithm of the characteristic function of $\{X_t\}$, which is

$$g_X(t) = i \mu' t - \frac{1}{2} t' \sum_\text{cumulant} t.$$  

Since the spectral density is the Fourier transform of the covariance it is natural to ask whether one can define the higher order spectral density as the fourier transform of the higher order cumulants. This turns out to be the case, and the higher order spectra have several interesting properties.

Let us suppose that $\{X_t\}$ is a stationary time series (notice that we are assuming it is strictly stationary and not second order). Let $\kappa_3(t, s) = \text{cum}(X_0, X_t, X_s)$, $\kappa_3(t, s, r) = \text{cum}(X_0, X_t, X_s, X_r)$ and $\kappa_q(t_1, \ldots, t_{q-1}) = \text{cum}(X_0, X_{t_1}, \ldots, X_{t_q})$ (noting that like the covariance the higher order cumulants are invariant to shift). The third, fourth and the general $q^{th}$ order spectra is defined as

$$f_3(\omega_1, \omega_2) = \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \kappa_3(s, t) \exp(is\omega_1 + it\omega_2)$$

$$f_4(\omega_1, \omega_2, \omega_3) = \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \kappa_4(s, t, r) \exp(is\omega_1 + it\omega_2 + ir\omega_3)$$

$$f_q(\omega_1, \omega_2, \ldots, \omega_{q-1}) = \sum_{t_1, \ldots, t_{q-1}=-\infty}^{\infty} \kappa_q(t_1, t_2, \ldots, t_{q-1}) \exp(it_1\omega_1 + it_2\omega_2 + \ldots + it_{q-1}\omega_{q-1}).$$

Example 8.6.1 (Third and Fourth order spectral density of a linear process) Let us sup-
pose that \{X_t\} satisfies

\[ X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \]

where \( \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \), \( \text{E}(\varepsilon_t) = 0 \) and \( \text{E}(\varepsilon_t^4) < \infty \). Let \( A(\omega) = \sum_{j=-\infty}^{\infty} \psi_j \exp(ij\omega) \). Then it is straightforward to show that

\[ f(\omega) = \sigma^2 |A(\omega)|^2 \]
\[ f_3(\omega_1, \omega_2) = \kappa_3 A(\omega_1)A(\omega_2)A(-\omega_1 - \omega_2) \]
\[ f_4(\omega_1, \omega_2, \omega_3) = \kappa_4 A(\omega_1)A(\omega_2)A(\omega_3)A(-\omega_1 - \omega_2 - \omega_3), \]

where \( \kappa_3 = \text{cum}(\varepsilon_t, \varepsilon_t, \varepsilon_t) \) and \( \kappa_4 = \text{cum}(\varepsilon_t, \varepsilon_t, \varepsilon_t, \varepsilon_t) \).

We see from the example, that unlike the spectral density, the higher order spectras are not necessarily positive or even real.

A review of higher order spectra can be found in Brillinger (2001). Higher order spectras have several applications especially in nonlinear processes, see Subba Rao and Gabr (1984). We will consider one such application in a later chapter.

Using the definition of the higher order spectrum we can now generalise Lemma 8.2.1 to higher order cumulants (see Brillinger (2001), Theorem 4.3.4).

**Proposition 8.6.1** \{X_t\} is a strictly stationary time series, where for all \( 1 \leq i \leq q - 1 \) we have \( \sum_{t_1, \ldots, t_{q-1}=\infty}^{\infty} |(1 + t_i)\kappa_q(t_1, \ldots, t_{q-1})| < \infty \) (note that this is simply a generalization of the covariance assumption \( \sum_{r} |rc(r)| < \infty \)). Then we have

\[
\text{cum}(J_n(\omega_{k_1}), \ldots, J_n(\omega_{k_q})) = \frac{1}{n^{q/2}} f_q(\omega_{k_2}, \ldots, \omega_{k_q}) \sum_{j=1}^{n} \exp(ij(\omega_{k_1} - \ldots - \omega_{k_q})) + O\left(\frac{1}{n^{q/2}}\right)
\]
\[
= \begin{cases} 
\frac{1}{n^{(q-1)/2}} f_q(\omega_{k_2}, \ldots, \omega_{k_q}) + O\left(\frac{1}{n^{q/2}}\right) & \sum_{i=1}^{q} k_i = nZ \\
O\left(\frac{1}{n^{q/2}}\right) & \text{otherwise}
\end{cases}
\]

where \( \omega_{k_i} = \frac{2\pi k_i}{n} \).
8.7 Extensions

8.7.1 The spectral density of a time series with randomly missing observations

Let us suppose that \( \{X_t\} \) is a second order stationary time series. However \( \{X_t\} \) is not observed at everytime point and there are observations missing, thus we only observe \( X_t \) at \( \{\tau_k\}_k \). Thus what is observed is \( \{X_{\tau_k}\} \). The question is how to deal with this type of data. One method was suggested in \cite{?}. He suggested that the missingness mechanism \( \{\tau_k\} \) be modelled stochastically. That is define the random process \( \{Y_t\} \) which only takes the values \( \{0, 1\} \), where \( Y_t = 1 \) if \( X_t \) is observed, but \( Y_t = 0 \) if \( X_t \) is not observed. Thus we observe \( \{X_t Y_t\}_t = \{X_{\tau_k}\} \) and also \( \{Y_t\} \) (which is the time points the process is observed). He also suggests modelling \( \{Y_t\} \) as a stationary process, which is independent of \( \{X_t\} \) (thus the missingness mechanism and the time series are independent).

The spectral densities of \( \{X_t Y_t\} \), \( \{X_t\} \) and \( \{Y_t\} \) have an interest relationship, which can be exploited to estimate the spectral density of \( \{X_t\} \) given estimators of the spectral densities of \( \{X_t Y_t\} \) and \( \{X_t\} \) (which we recall are observed). We first note that since \( \{X_t\} \) and \( \{Y_t\} \) are stationary, then \( \{X_t Y_t\} \) is stationary, furthermore

\[
\text{cov}(X_t Y_t, X_r Y_r) = \text{cov}(X_t, X_r) \text{cov}(Y_t, Y_r) + \text{cov}(X_t, Y_r) \text{cov}(Y_t, X_r) + \text{cum}(X_t, Y_t, X_r, Y_r) \\
= \text{cov}(X_t, X_r) \text{cov}(Y_t, Y_r) = c_X(t - \tau)c_Y(t - \tau)
\]

where the above is due to independence of \( \{X_t\} \) and \( \{Y_t\} \). Thus the spectral density of \( \{X_t Y_t\} \) is

\[
f_{XY}(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \text{cov}(X_0 Y_0, X_r Y_r) \exp(ir\omega) \\
= \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c_X(r)c_Y(r) \exp(ir\omega) \\
= \int f_X(\lambda)f_Y(\omega - \lambda) d\omega,
\]

where \( f_X(\lambda) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c_X(r) \exp(ir\omega) \) and \( f_Y(\lambda) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c_Y(r) \exp(ir\omega) \) are the spectral densities of the observations and the missing process.