Chapter 8

Spectral Analysis

8.1 Some Fourier background

The background given here is extremely sketchy (to say the least), for a more thorough background the reader is referred, for example, to Priestley (1983), Chapter 4 and Fuller (1995), Chapter 3.

(i) Fourier transforms of finite sequences

It is straightforward to show (by using that $\sum_{j=1}^{n} \exp(i2\pi k/n) = 0$ for $k \neq 0$) that if

$$d_k = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} x_j \exp(i2\pi jk/n),$$

then $\{x_r\}$ can be recovered by inverting this transformation

$$x_r = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} d_k \exp(-i2\pi rk/n),$$

(ii) Fourier sums and integrals

Of course the above only has meaning when $\{x_k\}$ is a finite sequence. However suppose that $\{x_k\}$ is a sequence which belongs to $\ell_2$ (that is $\sum_{k} x_k^2 < \infty$), then we can define the function

$$f(\omega) = \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} x_k \exp(ik\omega),$$

where $\int_{0}^{2\pi} f(\omega)^2 d\omega = \sum_{k} x_k^2$, and we can recover $\{x_k\}$ from $f(\omega)$. That is

$$x_k = \frac{1}{\sqrt{2\pi}} \int_{0}^{2\pi} f(\omega) \exp(-ik\omega).$$
(iii) **Convolutions.** Let us suppose that the Fourier transform of the sequence \( \{a_k\} \) is \( A(\omega) = \frac{1}{\sqrt{2\pi}} \sum_k a_k \exp(ik\omega) \) and Fourier transform of the sequence \( \{b_k\} \) is \( B(\omega) = \frac{1}{\sqrt{2\pi}} \sum_k b_k \exp(ik\omega) \). Then

\[
\sum_{j=-\infty}^{\infty} a_j b_{k-j} = \int A(\omega)B(-\omega)\exp(-ik\omega) d\omega \\
\sum_{j=-\infty}^{\infty} a_j b_j \exp(ij\omega) = \int A(\lambda)B(\omega - \lambda) d\lambda. \tag{8.1}
\]

### 8.2 Motivation

To give a taster of the spectral representations below let us consider the following example.

Suppose that \( \{X_t\}_{t=1}^{n} \) is a stationary time series. The Fourier transform of this sequence is

\[
J_n(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \exp(it\omega_j)
\]

where \( \omega_j = 2\pi j/n \) (these are often called the fundamental frequencies). Using (i) above we see that

\[
X_t = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} J_n(\omega_j) \exp(-ik\omega_t). \tag{8.2}
\]

This is just the inverse Fourier transform, however \( \{J_n(\omega_j)\} \) has some interesting properties. Under certain conditions it can be shown that \( \operatorname{cov}(J_n(\omega_s), J_n(\omega_t)) \approx 0 \) if \( s \neq t \). So in some sense (8.2) can be considered as the decomposition of \( X_t \) in terms of frequencies whose amplitudes are uncorrelated. Now if we let \( f_n(\omega) = n^{-1}E(|J_n(\omega)|^2) \), and take the above argument further we see that

\[
c(k) = \operatorname{cov}(X_k, X_{k+t}) = \frac{1}{n} \sum_{s=1}^{n} E(|J_n(\omega_s)|^2) \exp(ik\omega_s - i(k + t)\omega_s) = \sum_{s=1}^{n} f_n(\omega_s) \exp(-ik\omega_s). \tag{8.3}
\]

For more details on this see Priestley (1983), Section 4.11 (pages 259-261). Note that the above can be considered as the eigen decomposition of the stationary covariance function, since

\[
c(u, v) = c(u - v) = \sum_{s=1}^{n} f_n(\omega_s) \exp(iu\omega_s) \exp(-iv\omega_s),
\]

where \( \{\exp(it\omega_s)\} \) are the eigenfunctions and \( f_n(\omega_s) \) the eigenvalues.

Of course the entire time series \( \{X_t\} \) will have infinite length (and in general it will not belong to \( \ell_2 \)), so it is natural to ask whether the above results can be generalised to \( \{X_t\} \). The answer is yes, by replacing the sum in (8.3) by an integral to obtain

\[
c(k) = \int_{0}^{2\pi} \exp(ik\omega) dF(\omega),
\]

89
where $F(\omega)$ is a positive nondecreasing function. Comparing with (8.3), we observe that $f_n(\omega_k)$ is a positive function, thus its integral (the equivalent of $F(\omega)$) is positive and nondecreasing. Therefore heuristically we can suppose that $F(\omega) \approx \int_0^\omega f_n(\lambda) d\lambda$.

Moreover the analogue of (8.2) is

$$X_t = \int \exp(ik\omega) dZ(\omega),$$

where $Z(\omega)$ is right continuous orthogonal increment process (that is $\mathbb{E}((Z(\omega_1) - Z(\omega_2))(Z(\omega_3) - Z(\omega_4)) = 0$, when the intervals $[\omega_1, \omega_2]$ and $[\omega_3, \omega_4]$ do not overlap) and $\mathbb{E}(|Z(\omega)|^2) = F(\omega)$.

We mention that a more detailed discussion on spectral analysis in time series is given 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 Spectral representations

#### 8.3.1 The spectral distribution

We first state a theorem which is very useful for checking positive definiteness of a sequence. See Brockwell and Davis (1998), Corollary 4.3.2 or Fuller (1995), Theorem 3.1.9.

To prove part of the result we use the fact that if a sequence $\{a_k\} \in \ell_2$, then $g(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} a_k \exp(ik\omega) \in L_2$ (by Parseval’s theorem) and $a_k = \int_0^{2\pi} g(\omega) \exp(ik\omega)$ and the following result.

**Lemma 8.3.1** Suppose $\sum_{k=-\infty}^{\infty} |c(k)| < \infty$, then we have

$$\frac{1}{n} \sum_{k=-(n-1)}^{(n-1)} |kc(k)| \to 0$$

as $n \to \infty$.

**PROOF.** The proof is straightforward in the case that $\sum_{k=-\infty}^{\infty} |kc(k)| < \infty$, in this case $\sum_{k=-(n-1)}^{(n-1)} \frac{|k|}{n} |c(k)| = O\left(\frac{1}{n}\right)$. The proof is slightly more tricky in the case that $\sum_{k=-\infty}^{\infty} |c(k)| < \infty$. First we note that since $\sum_{k=-\infty}^{\infty} |c(k)| < \infty$ for every $\varepsilon > 0$ there exists a $N_\varepsilon$ such that for all $n \geq N_\varepsilon$, $\sum_{|k| \geq n} |c(k)| < \varepsilon$. Let us suppose that $n > N_\varepsilon$, then we have the bound

$$\frac{1}{n} \sum_{k=-(n-1)}^{(n-1)} |kc(k)| \leq \frac{1}{n} \sum_{k=-(N_\varepsilon-1)}^{(N_\varepsilon-1)} |kc(k)| + \frac{1}{n} \sum_{|k| \geq N_\varepsilon} |kc(k)|$$

$$\leq \frac{1}{2\pi n} \sum_{k=-(N_\varepsilon-1)}^{(N_\varepsilon-1)} |kc(k)| + \varepsilon.$$ 

Hence if we keep $N_\varepsilon$ fixed we see that $\frac{1}{n} \sum_{k=-(N_\varepsilon-1)}^{(N_\varepsilon-1)} |kc(k)| \to 0$ as $n \to \infty$. Since this is true for all $\varepsilon$ (for different thresholds $N_\varepsilon$) we obtain the required result. \qed
Theorem 8.3.1 (The spectral density) Suppose the coefficients \( \{c(k)\} \) are absolutely summable (that is, \( \sum_k |c(k)| < \infty \)). Then the sequence \( \{c(k)\} \) is nonnegative 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

\[
c(k) = \int_0^{2\pi} \exp(ik\omega)f(\omega) d\omega. \tag{8.4}
\]

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 non-negative 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

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

Clearly \( f_n(\omega) \geq 0 \). We note that \( f_n(\omega) \) can be rewritten as

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

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

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

\[
= I_n + II_n.
\]

Now since \( \sum_{k=-\infty}^{\infty} |c(k)| < \infty \) it is clear that \( I_n \rightarrow 0 \) as \( n \rightarrow \infty \). Using Lemma 8.3.1 we have \( II_n \rightarrow 0 \) as \( n \rightarrow \infty \). Altogether the above implies

\[
|f(\omega) - f_n(\omega)| \rightarrow 0 \quad n \rightarrow \infty. \tag{8.5}
\]

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 that \( f(\omega) \), defined by \( f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} c(k) \exp(ik\omega) \), is a nonnegative function then \( \{c(k)\} \) is a nonnegative sequence. We first note because \( \{c(k)\} \in \ell_1 \) it is also in \( \ell_2 \) hence we have that \( c(k) = \int_{0}^{2\pi} f(\omega) \exp(ik\omega) \). Now we have

\[
\sum_{s,t=1}^{n} x_s c(s-t) \bar{x}_s = \int_{0}^{2\pi} f(\omega) \left\{ \sum_{s,t=1}^{n} x_s \exp(i(s-t)\omega) \bar{x}_s \right\} d\omega = \int_{0}^{2\pi} f(\omega) \left\{ \sum_{s=1}^{n} x_s \exp(is\omega) \right\}^2 d\omega \geq 0.
\]
Hence we obtain the desired result.

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 1.1.1).

**Example 8.3.1** Suppose we define the empirical covariances

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

then \( \{\hat{c}_n(k)\} \) is positive definite sequence. Therefore, using Lemma 1.1.1 there exists a stationary time series \( \{Z_t\} \) which has the covariance \( \hat{c}_n(k) \).

To show that the sequence is non-negative definite we will consider the Fourier transform of the sequence (the spectral density) and show that it is nonnegative. The Fourier transform of \( \{\hat{c}_n(k)\} \) is

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

Since it is positive, this means that \( \{\hat{c}_n(k)\} \) is a positive definite sequence.

We now state a useful result which relates the largest and smallest eigenvalue of of a variance matrix of a stationary process to the smallest and largest values of the spectral density.

**Lemma 8.3.2** Suppose that \( \{X_k\} \) is a stationary process with covariance function \( \{c(k)\} \) and spectral density \( f(\omega) \). Let \( \Sigma_n = \text{var}(X_n) \), where \( X_n = (X_1, \ldots, X_n) \). Suppose \( \inf_{\omega} f(\omega) \geq m > 0 \) and \( \sup_{\omega} f(\omega) \leq M < \infty \). Then for all \( n \) we have

\[
\lambda_{\min}(\Sigma_n) \geq \inf_{\omega} f(\omega) \quad \lambda_{\max}(\Sigma_n) \leq \sup_{\omega} f(\omega).
\]

**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

\[
\lambda_{\min}(\Sigma_n) = \langle \xi_1, \Sigma_n \xi_1 \rangle = \sum_{s,t=1}^{n} \overline{\xi}_{s,1} c(s-t) \xi_{t,1} = \int f(\omega) \sum_{s,t=1}^{n} \overline{e}_{s,1} \exp(i(s-t)\omega) e_{t,1} d\omega = \\
\int_0^{2\pi} f(\omega) \left| \sum_{s=1}^{n} e_{s,1} \exp(is\omega) \right|^2 d\omega \geq \int_0^{2\pi} f(\omega) \int_0^{2\pi} \left| \sum_{s=1}^{n} e_{s,1} \exp(is\omega) \right|^2 d\omega \geq \inf_{\omega} f(\omega),
\]

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

A consequence of the above result is if a spectral density is bounded from above and bounded away from zero then it is non-singular and with a bounded spectral norm.

**Lemma 8.3.3** 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 specify that the covariances are absolutely summable).
PROOF. See Brockwell and Davis (1998), Proposition 5.1.1. □

Theorem 8.3.1 only holds when the sequence \( \{c(k)\} \) is absolutely summable. Of course this may not always be the case. An example of an 'extreme' case is the time series \( X_t = Z \). Clearly this is a stationary time series and its covariance is \( c(k) = \text{var}(Z) \) for all \( k \). In this case the autocovariances \( \{c(k) = 1\} \), is not absolutely summable, hence the representation of the covariance in Theorem 8.3.1 can not be applied to this case. The reason is because the fourier transform of the infinite sequence \( \{1\} \) is not well defined (since \( \{1\} \) does not belong to \( \ell_1 \) and also \( \ell_2 \)).

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 (we use the integral \( \int g(x) dF(x) \), a definition is given in the Appendix).

**Theorem 8.3.2** A function \( \{c(k)\} \) is non-negative definite sequence if and only if

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

(8.6)

where \( F(\omega) \) is a right-continuous (this means that \( f(x + h) \rightarrow f(x) \) as \( 0 < h \rightarrow 0 \), 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.

PROOF. 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. Had \( \{c(k)\} \) been 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 \( f(\lambda) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} c(k) \exp(ik\omega) \). By using Theorem 8.3.1 we know that \( f(\lambda) \) is nonnegative, hence \( F(\omega) \) is a distribution, and we have the result.

In the case that \( \{c(k)\} \) is not absolutely summable we cannot use this approach but 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

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

When the \( \{c(k)\} \) is not absolutely summable, the limit of \( f_n(\omega) \) may no longer be well defined. To circumvent our dealing with functions which may have awkward limits, we consider instead their integral, which we will show will always be a distribution function. Let us define the function \( F_n(\omega) \) whose derivative is \( f_n(\omega) \), that is

\[
F_n(\omega) = \int_0^\omega f_n(\lambda) d\lambda \quad 0 \leq \lambda \leq 2\pi.
\]

Since \( f_n(\lambda) \) is nonnegative, \( F_n(\omega) \) is a nondecreasing function, and it is bounded \( (F_n(\pi) = \int_0^{2\pi} f_n(\lambda) d\lambda \leq c(0)) \). Hence \( F_n \) satisfies all properties of a distribution and can be treated as a distribution function. Now it is clear that for every \( k \) we have

\[
\int_0^{2\pi} \exp(ik\omega) dF_n(\omega) = \left\{ \begin{array}{ll} (1 - \frac{|k|}{n})c(k) & |k| \leq n \\ 0 & 0 \end{array} \right. \quad \quad \quad (8.7)
\]
If we let $d_{n,k} = \int_0^{2\pi} \exp(ik\omega) dF_n(\omega)$, we see that for every $k$, $d_{n,k} \rightarrow d_k$ as $n \rightarrow \infty$. But we should ask what this tells us about the limit of the distribution $\{F_n\}$? Intuitively, the distributions $\{F_n\}$ should (weakly converge) to a function $F$ and this function should also be a distribution function (if $\{F_n\}$ are all nondecreasing functions, then its limit must be nondecreasing). In fact this turns out to be the case by applying Helly’s theorem (see Appendix). Roughly speaking, it states that given a sequence of distributions $\{G_k\}$ which are all bounded, then there exists a distribution $G$, which is the limit of a subsequence of $\{G_k\}$ (effectively this determining conditions for a sequence of functions to be compact), hence for every $h \in L_2$ we have

$$\int h(\omega) dG_{k_i}(\omega) \rightarrow \int h(\omega) dG(\omega).$$

We now apply this result to the sequence $\{F_n\}$. We observe that the sequence of distributions $\{F_n\}$ are all uniformly bounded (by $c(0)$). Therefore applying Helly’s theorem there must exist a distribution $F$, which is the limit of a subsequence of $\{F_n\}$, that is for every $h \in L_2$ we have

$$\int h(\omega) dF_{k_i}(\omega) \rightarrow \int h(\omega) dF(\omega), \quad i \rightarrow \infty,$$

for some subsequence $\{F_{k_i}\}$. We now show that above is true not only for a subsequence but the actual sequence $\{F_k\}$. We observe that $\{\exp(ik\omega)\}$ is a basis of $L_2$, and that the sequence $\{\int_0^{2\pi} \exp(ik\omega) dF_n(\omega)\}_n$ converges for all $k$, to $c(k)$. Therefore for all $h \in L_2$ we have

$$\int h(\omega) dF_k(\omega) \rightarrow \int h(\omega) dF(\omega), \quad k \rightarrow \infty,$$

for some distribution function $F$. Therefore looking at the case $\{\exp(ik\omega)\}$ we have

$$\int \exp(ik\omega) dF_k(\omega) \rightarrow \int \exp(ik\omega) dF(\omega), \quad k \rightarrow \infty.$$

Since $\int \exp(ik\omega) dF_k(\omega) \rightarrow c(k)$ and $\int \exp(ik\omega) dF_k(\omega) \rightarrow \int \exp(ik\omega) dF(\omega)$, then we have $c(k) = \int \exp(ik\omega) dF(\omega)$, where $F$ is a distribution.

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

**Example 8.3.2** We now construct the spectral distribution for the 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

$$\text{cov}(X_0, X_k) = \text{var}(Z) = \int \exp(ik\omega) dF(\omega).$$

**8.3.2 The spectral representation theorem**

We now state the spectral representation theorem and give a rough outline of the proof.
Remark 8.3.1 Note that so far we have not defined the integral on the right hand side of (8.8), this is known as a stochastic integral. Unlike many deterministic functions (functions whose derivative exists), one cannot really suppose (8.8), this is known as a stochastic integral. Unlike many deterministic functions (functions $f \in H$), Brownian is quite ‘rough’, that is a typical realisation of Brownian motion satisfies

$$\exp(\omega) \in \mathbb{F}^2$$

where for $\omega_1 \geq \omega_2$, $\mathbb{E}(Z(\omega_1) - Z(\omega_2))^2 = F(\omega_1) - F(\omega_2)$ (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 (8.8) 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 8.3.1 Note that so far we have not defined the integral on the right hand side of (8.8), this 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. In other words let $Z_n(\omega) = \sum_{k=1}^{n} Z(\omega_k)I_{\omega_{k-1}, \omega_k}(\omega)$ and

$$\int g(\omega)dZ_n(\omega) = \sum_{k=1}^{n} g(\omega_k)(Z(\omega_k) - Z(\omega_{k-1})),$$

then $\int g(\omega)dZ(\omega)$ is the mean squared limit of $\{\int g(\omega)dZ_n(\omega)\}_n$ that is $\mathbb{E}[\int g(\omega)dZ(\omega) - \int g(\omega)dZ_n(\omega)]^2$.

For a more precise explanation, see Priestley (1983), Sections 3.6.3 and Section 4.11 and Brockwell and Davis (1998), Section 4.7.

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.

**PROOF of the Spectral Representation Theorem** To prove the result we will 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. We will define what is known as an isomorphism (a one-to-one mapping which preserves the norm and is linear) between these two spaces.

Let $\mathcal{H}_1$ be defined by all functions $f$, if $\int_0^{2\pi} f^2(\omega)dF(\omega) < \infty$, then $f \in \mathcal{H}_1$ and define the inner product on $\mathcal{H}_1$ to be

$$\langle f, g \rangle = \int_0^{2\pi} f(x)g(x)dF(x). \quad (8.9)$$

We first note that $\{\exp(ik\omega)\}$ belongs to $\mathcal{H}_1$, moreover they also span the space $\mathcal{H}_1$. Hence if $f \in \mathcal{H}_1$, then there exists coefficients $\{a_j\}$ such that $f(x) = \sum_j a_j \exp(ij\omega)$. Let $\mathcal{H}_2$ be the
space spanned by \( \{X_i\} \), hence \( \mathcal{H}_2 = \bar{s}\mathcal{P}(\{X_i\}) \) (it necessary to define the closure of this space, but we won’t do so here) and the inner product is the covariance \( \text{cov}(Y, X) \).

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

\[
T(\sum_{j=1}^{n} a_j \exp(ik\omega)) = \sum_{j=1}^{n} a_j X_k,
\]

for any \( n \) (it is necessary to show that this can be extended to infinite \( n \), but we won’t do so here). We need to shown that \( T \) defines an isomorphism. We first observe that this mapping preserves the inner product. That is suppose \( f, g \in \mathcal{H}_1 \), then there exists \( \{f_j\} \) 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 (8.10) we have

\[
<Tf, Tg> = \text{cov}(\sum_j f_j X_j, \sum_j g_j X_j) = \sum_{j_1, j_2} f_{j_1} g_{j_2} \text{cov}(X_{j_1}, X_{j_2})
\]

\[
= \int_{0}^{2\pi} (\sum_{j_1, j_2} f_{j_1} g_{j_2} \exp(i(j_1 - j_2)\omega))dF(\omega) = \int_{0}^{2\pi} f(x)g(x)dF(x) = <f, g>.
\]

Hence \( <Tf, Tg> = <f, g> \), so the inner product is preserved. To show that it is a one-to-one mapping see Brockwell and Davis (1998), Section 4.7. 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 display many similar properties.

Since for all \( \omega \in [0, 2\pi] \), the identity functions \( I_{[0,\omega]}(x) \in \mathcal{H}_1 \), we can define the random function \( \{Z(\omega); 0 \leq \lambda \leq 2\pi\} \) to be \( T(I_{[0,\omega]}) = Z(\omega) \). Now since that mapping \( T \) is linear we observe that \( T(I_{[\omega_1, \omega_3]}) = Z(\omega_1) - Z(\omega_2) \). Moreover, since \( T \) preserves the norm we have for any non-intersecting intervals \([\omega_1, \omega_2]\) and \([\omega_3, \omega_4]\) that

\[
\mathbb{E}((Z(\omega_1) - Z(\omega_2))(Z(\omega_3) - Z(\omega_4))) = <T(I_{[\omega_1, \omega_2]}), T(I_{[\omega_3, \omega_4]})> = <I_{[\omega_1, \omega_2]}, I_{[\omega_3, \omega_4]}> = \int I_{[\omega_1, \omega_2]} \omega \omega dF(\omega) = 0.
\]

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

\[
\mathbb{E}((Z(\omega_1) - Z(\omega_2))^2) = <T(I_{[\omega_1, \omega_2]}), T(I_{[\omega_1, \omega_2]})> = <I_{[\omega_1, \omega_2]}, I_{[\omega_1, \omega_2]}> = \int_{\omega_1}^{\omega_2} dF(\omega) = F(\omega_1) - F(\omega_2).
\]

Having defined the two spaces which are isomorphic and the random function \( \{Z(\omega); 0 \leq \lambda \leq 2\pi\} \) and function \( I_{[0,\omega]}(x) \) which are have orthogonal increments. We can now prove the result. We note that for any function \( g \in L_2 \) we can write

\[
g(\omega) = \int_{0}^{2\pi} g(s) dI(s - \omega),
\]

where \( I(s) \) is the identity function with \( I(s) = 0 \), for \( s < 0 \), and \( I(s) = 1 \), for \( s \geq 0 \) (hence \( dI(\omega - s) = \delta_\omega(s)ds \) and \( \delta_\omega(s) \) is the dirac delta function). We now consider the special case \( g(t) = \exp(it\omega) \), and apply the isomophism \( T \) to this

\[
T(\exp(it\omega)) = \int_{0}^{2\pi} \exp(its) dT(I(\omega - s)),
\]

96
where the mapping goes inside the integral due to the linearity of the isomorphism. Now we observe that
\[ I(s - \omega) = I[0,s](\omega) \]
and by definition of \( \{Z(\omega); 0 \leq \lambda \leq 2\pi\} \) we have \( T(I[0,s](\omega)) = Z(s) \). Substituting this into the above gives
\[ X_t = \int_0^{2\pi} \exp(its)dZ(s), \]
which gives the required result. \( \square \)

It is worth pointing out in the above proof, the exponential functions \( \{\exp(ik\omega)\} \) do not necessarily play a unique role. Indeed 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
\[ \mathbb{E}(X_tX_\tau) = c(t, \tau) = \int_0^{2\pi} \phi_t(\omega)\phi_\tau(\omega)dH(\omega). \]

So long as this representation exists then we can define two spaces \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) where \( \{\phi_k\} \) is the basis of the functional space \( \mathcal{H}_1 \) and it contains all functions \( f \) such that \( \int f^2(\omega)dH(\omega) < \infty \) and \( \mathcal{H}_2 \) is the random space defined by \( \tilde{\phi}(\ldots, X_{-1}, X_0, X_1, \ldots) \). Now we can define an isomorphism \( T: \mathcal{H}_1 \rightarrow \mathcal{H}_2 \), where for all functions \( f(\omega) = \sum_k f_k \phi_k(\omega) \in \mathcal{H}_1 \)
\[ T(f) = \sum_k f_k X_k \in \mathcal{H}_2. \]

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 \( \mathbb{E}|Z(\omega)|^2 = H(\omega) \). We state this result in the theorem below (see Priestley (1983), Section 4.11).

**Theorem 8.3.4 (General orthogonal expansions)** Let \( \{X_t\} \) be a time series (not necessarily second order stationary) with covariance \( \{\mathbb{E}(X_tX_\tau) = c(t, s)\} \). If there exists a sequence of functions \( \{\phi_k(\cdot)\} \) which satisfy for all \( k \)
\[ \int_0^{2\pi} |\phi_k(\omega)|^2dH(\omega) \]
and the covariance admits the representation
\[ c(t, s) = \int_0^{2\pi} \phi_t(\omega)\bar{\phi}_s(\omega)dH(\omega), \quad (8.11) \]
where \( H \) is a distribution then for all \( t \) we have the representation
\[ X_t = \int \phi_t(\omega)dZ(\omega), \quad (8.12) \]
where \( \{Z(\omega)\} \) are orthogonal random functions and \( \mathbb{E}|Z(\omega)|^2 = H(\omega) \). On the other hand if \( X_t \) has the representation \( (8.12) \), then \( c(s, t) \) admits the representation \( (8.11) \).
Remark 8.3.2 We mention that the above representation applies to both stationary and non-stationary 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:

\[
c(k) = \int_{0}^{2\pi} \exp(ik\omega)dF(\omega).
\] (8.13)

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.

We mention that it is not always guaranteed that for any basis \( \{\phi_t\} \) we can represent the covariance \( \{c(k)\} \) as (8.11). However (8.12) is a very useful starting point for characterising nonstationary processes.

8.3.3 The spectral densities 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

\[
X_t = \sum_{j=\infty}^{\infty} \psi_j \varepsilon_{t-j}
\] (8.14)

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

\[
c(k) = \text{cov}(X_t, X_{t+k}) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k}.
\] (8.15)

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 (8.15) 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 spectral representation. We prove the results using the latter method.

Since \( \{\varepsilon_t\} \) are iid random variables, using Theorem 8.3.3 there exists an orthogonal random function \( \{Z(\omega)\} \) such that

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

Using the above we can obtain the spectral representation for $\{X_t\}$

$$X_t = \frac{1}{2\pi} \int_0^{2\pi} \left\{ \sum_{j=-\infty}^{\infty} \psi_j \exp(-ij\omega) \right\} \exp(it\omega) d\omega.$$

Hence

$$X_t = \int_0^{2\pi} A(\omega) \exp(it\omega) d\omega,$$

where $A(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \psi_j \exp(-ij\omega)$, noting that this is the unique spectral representation of $X_t$.

Now multiplying the above by $X_{t+k}$ and taking expectations gives

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

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

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

Comparing the above with (8.4) we see that the spectral density $f(\omega) = 2\pi \sigma^2 |A(\omega)|^2 = \frac{\sigma^2}{2\pi} \sum_{j=-\infty}^{\infty} \psi_j \exp(-ij\omega)^2$. Therefore the spectral density function corresponding to the MA($\infty$) process defined in (8.14) is

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

**Example 8.3.3** 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 $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t^2) = \sigma^2$. Then the spectral density of $\{X_t\}$ is

$$f(\omega) = \frac{\sigma^2 |1 + \sum_{j=1}^{q} \theta_j \exp(ij\omega)|^2}{2\pi |1 - \sum_{j=1}^{\infty} \phi_j \exp(ij\omega)|^2}$$

We note that because the ARMA is the ratio of trigonometric polynomials, this is known as a rational spectral density.
Definition 8.3.1 (The Cramer Representation) We mention that the representation in (8.16) 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.

8.3.4 Higher order spectrums

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

\[ \text{cum}(X_1, X_2, X_3) = \mathbb{E}(X_1X_2X_3) \]

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

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

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 moment generating function of \( \{X_t\} \).

Since the spectral density is the fourier transform of the covariance it is natural to ask whether one can define the higher order spectra 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 \( \text{cum}(t, s) = \mathbb{E}(X_0, X_t, X_s) \) and \( \text{cum}(t, s, r) = \mathbb{E}(X_0, X_t, X_s, X_r) \) (noting that like the covariance the higher order cumulants are invariant to shift). The third and fourth order spectras is defined as

\[
\begin{align*}
  f(\omega_1, \omega_2) &= \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \text{cum}(s, t) \exp(is\omega_1 + it\omega_2) \\
  f(\omega_1, \omega_2, \omega_3) &= \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \text{cum}(s, t, r) \exp(is\omega_1 + it\omega_2 + ir\omega_3). 
\end{align*}
\]

Example 8.3.4 (Third and Fourth order spectra of a linear process) Let us suppose that \( \{X_t\} \) satisfies

\[ X_t = \sum_{j=-\infty}^{\infty} \psi_j \xi_{t-j} \]
where \( \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \), \( \mathbb{E}(\varepsilon_t) = 0 \) and \( \mathbb{E}(\varepsilon_t^2) < \infty \). Let \( A(\omega) = \sum_{j=-\infty}^{\infty} \psi_j \exp(ij\omega) \). Then it is straightforward to show that

\[
\begin{align*}
    f_3(\omega_1, \omega_2) &= \kappa_3 A(\omega_1) A(\omega_2) A(-\omega_1 - \omega_2) \\
    f_4(\omega_1, \omega_2, \omega_3) &= \kappa_4 A(\omega_1) A(\omega_2) A(\omega_3) A(-\omega_1 - \omega_2 - \omega_3),
\end{align*}
\]

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 spectra are not necessarily positive or even real.

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

### 8.4 The Periodogram and the spectral density function

Our aim is to construct an estimator of the spectral density function \( f(\omega) \) associated with the second order stationary process \( \{X_t\} \).

#### 8.4.1 The periodogram and its properties

Let us suppose we observe \( \{X_t\}_{t=1}^{n} \) which are observations from a zero mean, second order stationary time series. Let us suppose that the autocovariance function is \( \{c(k)\} \), where \( c(k) = \mathbb{E}(X_t X_{t+k}) \) and \( \sum_k |c(k)| < \infty \). We recall that we can be estimate \( c(k) \) using

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

Given that the spectral density is

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

then a natural estimator of \( f(\omega) \) is

\[
I_X(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \hat{c}_n(k) \exp(ik\omega). \tag{8.17}
\]

We usually call \( I_X(\omega) \) the periodogram. We will show that the periodogram has several nice properties that make it a suitable candidate for the spectral density estimator. The only problem is that the raw periodogram turns out to be an inconsistent estimator. However with some modifications of the periodogram we can construct a good estimator of the spectral density.

**Lemma 8.4.1** Suppose that \( \{X_t\} \) is a second order stationary time series with \( \sum_k |c(k)| < \infty \) and \( I_X(\omega) \) is defined in (8.17). Then we have

\[
I_X(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \hat{c}_n(k) \exp(ik\omega) = \frac{1}{2\pi n} \sum_{t=1}^{n} X_t \exp(it\omega) \tag{8.18}
\]

101
\[ |\mathbb{E}(I_X(\omega)) - f(\omega)| \leq \frac{1}{2\pi} \left( \sum_{|k| \geq n} |c(k)| + \sum_{|k| \leq n} \frac{|k|}{n} |c(k)| \right) \to 0 \]  \hspace{1cm} (8.19)

as \( n \to \infty \). Hence in the case that \( \sum_{k=-\infty}^{\infty} |kc(k)| < \infty \) we have \( \mathbb{E}(I_X(\omega)) - f(\omega) = O\left(\frac{1}{n}\right) \). Moreover \( \text{var}\left(\frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \exp(it\omega)\right) = \mathbb{E}(I_X(\omega)) \).

**PROOF.** \( (8.18) \) follows immediately from the definition of the periodogram.

To obtain the first inequality in \( (8.19) \) is straightforward. To show that \( \sum_{|k| \geq n} |c(k)| + \sum_{|k| \leq n} \frac{|k|}{n} |c(k)| \to 0 \) as \( n \to \infty \), we note that \( \sum_{k} |c(k)| < \infty \), therefore \( \sum_{|k| \geq n} |c(k)| \to 0 \) as \( n \to \infty \). To show that \( \sum_{|k| \leq n} \frac{|k|}{n} |c(k)| \to 0 \) as \( n \to \infty \) we use \( \sum_{k} |c(k)| < \infty \) and Lemma 8.3.1, thus we obtain the desired result.

We see from the above that the periodogram is both a non-negative function as well as an asymptotically unbiased estimator of the spectral density. Hence is has inherited several of the characteristics of the spectral density. However a problem with the periodogram is that it is extremely erratic in its behaviour, in fact in its limit it does not converge to the spectral density. Hence as an estimator of the spectral density it is inappropriate. We will demonstrate this in the following two propositions and later discuss why this is so and how it can be made into a consistent estimator.

We start by considering the periodogram of iid random variables.

**Proposition 8.4.1** Suppose \( \{\varepsilon_t\}_{t=1}^{n} \) are iid random variables with mean zero and variance \( \sigma^2 \). We define \( J_\varepsilon(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \varepsilon_t \exp(it\omega) \) and \( I_\varepsilon(\omega) = \frac{1}{\sqrt{2\pi n}} \left| \sum_{t=1}^{n} \varepsilon_t \exp(it\omega) \right|^2 \). Then we have

\[ J_\varepsilon(\omega) \xrightarrow{D} \mathcal{N}(0, \frac{\sigma^2}{2\pi} I_2), \]  \hspace{1cm} (8.20)

for any finite \( m \)

\[ (J_\varepsilon(\omega_{k_1}), \ldots, J_\varepsilon(\omega_{k_m})) \xrightarrow{D} \mathcal{N}(0, \frac{\sigma^2}{2\pi} I_{2m}), \]  \hspace{1cm} (8.21)

\( I_\varepsilon(\omega)/\sigma^2 \sim \chi^2(2) \) (which is equivalent to the exponential distribution with mean one), \( (I_\varepsilon(\omega_{k_1}), \ldots, I_\varepsilon(\omega_{k_m})) \) converges in distribution to

\[ \text{cov}(I_\varepsilon(\omega_j), I_\varepsilon(\omega_k)) = \begin{cases} \frac{\kappa_4}{2\pi n} + \frac{2\kappa_2^2}{2\pi} & j \neq k \\ \frac{\kappa_4}{2\pi n} & j = k \end{cases} \]  \hspace{1cm} (8.22)

where \( \omega_j = 2\pi j/n \) and \( \omega_k = 2\pi k/n \) (and \( k, j \neq 0, n \)).

**PROOF.** We first show \( (8.20) \). We note that \( \Re(J_\varepsilon(\omega_k)) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \alpha_{t,n} \) and \( \Im(J_\varepsilon(\omega_k)) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \beta_{t,n} \) where \( \alpha_{t,n} = \varepsilon_t \cos(2k\pi t/n) \) and \( \beta_{t,n} = \varepsilon_t \sin(2k\pi t/n) \). We note that \( \Re(J_\varepsilon(\omega_k)) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \alpha_{t,n} \) and \( \Im(J_\varepsilon(\omega_k)) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \beta_{t,n} \) are the weighted sum of iid random variables, hence \( \{\alpha_{t,n}\} \) and \( \{\beta_{t,n}\} \) are martingale differences. Therefore, to show asymptotic normality, we will use the martingale central limit theorem with the Cramer-Wold device to show that
We show the result we need to verify the three conditions of the martingale CLT. First we consider the variances and the conditional variances

\[
\frac{1}{2\pi n} \sum_{t=1}^{n} \mathbb{E}(|\alpha_{t,n}|^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = \frac{1}{2\pi n} \sum_{t=1}^{n} \cos(2k\pi t/n)^2 \varepsilon_t^2 \xrightarrow{P} \frac{\sigma^2}{2\pi} \\
\frac{1}{2\pi n} \sum_{t=1}^{n} \mathbb{E}(|\beta_{t,n}|^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sin(2k\pi t/n)^2 \varepsilon_t^2 \xrightarrow{P} \frac{\sigma^2}{2\pi} \\
\frac{1}{2\pi n} \sum_{t=1}^{n} \mathbb{E}(\alpha_{t,n}\beta_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = \frac{1}{2\pi n} \sum_{t=1}^{n} \cos(2k\pi t/n) \sin(2k\pi t/n) \varepsilon_t^2 \xrightarrow{P} \mathbb{E}(\varepsilon_t^2) \frac{1}{\pi n} \sum_{t=1}^{n} \sin(2k \times 2\pi t/n) = 0.
\]

Finally we need to verify the Lindeberg condition, we only verify it for \( \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \alpha_{t,n} \). We note that for every \( \epsilon > 0 \) we have

\[
\frac{1}{2\pi n} \sum_{t=1}^{n} \mathbb{E}(|\alpha_{t,n}|^2 I(|\alpha_{t,n}| \geq 2\pi \sqrt{n}\epsilon) | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = \frac{1}{2\pi n} \sum_{t=1}^{n} \mathbb{E}(|\alpha_{t,n}|^2 I(|\alpha_{t,n}| \geq 2\pi \sqrt{n}\epsilon)) \\
\leq \frac{1}{2\pi n} \sum_{t=1}^{n} \mathbb{E}(|\varepsilon_t|^2 I(|\varepsilon_t| \geq 2\pi \sqrt{n}\epsilon)) = \mathbb{E}(|\varepsilon_t|^2 I(|\varepsilon_t| \geq 2\pi \sqrt{n}\epsilon)) \xrightarrow{P} 0
\]

as \( n \to \infty \), noting that the second to last inequality is because \( |\alpha_{t,n}| = |\cos(2\pi t/n)\varepsilon_t| \leq \varepsilon_t \). Hence we have verified Lindeberg condition and we obtain (8.20). The proof of (8.21) is similar, hence we omit the details. We can show that \( I_\varepsilon(\omega) \sim \chi^2(2) \), because \( I_\varepsilon(\omega) = \Re(J_\varepsilon(\omega))^2 + \Im(J_\varepsilon(\omega))^2 \), hence from (8.20) we have \( I_\varepsilon(\omega)/ \sim \chi^2(2) \).

To prove (8.22) we note that

\[
\text{cov}(I_\varepsilon(\omega_j), I_\varepsilon(\omega_k)) = \frac{1}{2\pi n^2} \sum_{k_1} \sum_{k_2} \sum_{t_1} \sum_{t_2} \text{cov}(X_{t_1}X_{t_1+k_1}, X_{t_2}X_{t_2+k_2}).
\]

We recall that

\[
\text{cov}(\varepsilon_{t_1}\varepsilon_{t_2+k_1}, \varepsilon_{t_2}\varepsilon_{t_2+k_2}) = \text{cov}(\varepsilon_{t_1}, \varepsilon_{t_2+k_1})\text{cov}(\varepsilon_{t_2}, \varepsilon_{t_2+k_2}) + \text{cov}(\varepsilon_{t_1}, \varepsilon_{t_2})\text{cov}(\varepsilon_{t_1+k_1}, \varepsilon_{t_2+k_2}) + \text{cum}(\varepsilon_{t_1}, \varepsilon_{t_1+k_1}, \varepsilon_{t_2}, \varepsilon_{t_2+k_2}).
\]

We note that since \{\varepsilon_t\} are iid random variables, then for most \( t_1, t_2, k_1 \) and \( k_2 \) the above covariance is zero. The exceptions are when \( t_1 = t_2 \) and \( k_1 = k_2 \) or \( t_1 = t_2 \) and \( k_1 = k_2 = 0 \) or \( t_1 - t_2 = k_1 = -k_2 \). Counting all these combinations we have

\[
\text{cov}(I_\varepsilon(\omega_j), I_\varepsilon(\omega_k)) = \frac{2}{2\pi n^2} \sum_{k_1} \sum_{t_1} \sum_{t_2} \exp(ik(\omega_j - \omega_k))\kappa_2^2 + \frac{1}{2\pi n^2} \sum_{t_1} \kappa_4
\]

where \( \kappa_2 = \text{var}(\varepsilon_t) \) and \( \kappa_4 = \text{cum}(\varepsilon_t, \varepsilon_t, \varepsilon_t, \varepsilon_t) \). We note that for \( j \neq k \), \( \sum_t \exp(ik(\omega_j - \omega_k)) = 0 \) and for \( j = k \), \( \sum_t \exp(ik(\omega_j - \omega_k)) = n \), substituting this into \( \text{cov}(I_\varepsilon(\omega_j), I_\varepsilon(\omega_k)) \) gives us the desired result. \( \square \)

We have seen that the periodogram for iid random variables does not converge to a constant and indeed its distribution is asymptotically exponential. This suggests that something similar
holds true for linear processes. This is the case. In the following lemma we show that the periodogram of a general linear process \( X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \) is

\[
I_X(\omega) = |\sum_j \psi_j \exp(\imath j \omega)|^2 I_\varepsilon(\omega) + o_p(1) = \frac{2\pi}{\sigma^2} f(\omega) I_\varepsilon(\omega) + o_p(1),
\]

where \( f(\omega) = \frac{\sigma^2}{2\pi} |\sum_j \psi_j \exp(i j \omega)|^2 \) is the spectral density of \( \{X_t\} \).

**Lemma 8.4.2** Let us suppose that \( \{X_t\} \) satisfy \( X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_t \), where \( \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \), and \( \{\varepsilon_t\} \) are iid random variables with mean zero and variance \( \sigma^2 \). Then we have

\[
J_X(\omega) = \{ \sum_j \psi_j \exp(i j \omega) \} I_\varepsilon(\omega) + Y_n(\omega), \tag{8.23}
\]

where \( Y_n(\omega) = \frac{1}{\sqrt{n}} \sum_j \psi_j \exp(i j \omega) U_{n,j} \), with \( U_{n,j} = \sum_{t=1-j}^{n-j} \exp(\imath t \omega) \varepsilon_t - \sum_{t=1}^{n} \exp(\imath t \omega) \varepsilon_t \), \( \mathbb{E}(Y_n(\omega))^2 \leq \left( \frac{1}{n^{1/2}} \sum_{j=-\infty}^{\infty} |\psi_j| \min(|j|, n)^{1/2} \right)^2 \). Furthermore

\[
I_X(\omega) = |\sum_j \psi_j \exp(i j \omega)|^2 |I_\varepsilon(\omega)|^2 + R_n(\omega), \tag{8.24}
\]

where \( \mathbb{E}(\sup_\omega |R_n(\omega)|) \to 0 \) and as \( n \to \infty \).

If in addition \( \mathbb{E}(\varepsilon_t^2) < \infty \) and \( \sum_{j=-\infty}^{\infty} |j^{1/2} \psi_j| < \infty \) then \( \mathbb{E}(\sup_\omega |R_n(\omega)|^2) = O(n^{-1}) \).

**PROOF.** See Priestley (1983), Theorem 6.2.1 or Brockwell and Davis (1998), Theorem 10.3.1. \( \Box \)

Using the above we see that \( I_X(\omega) \approx f(\omega) I_\varepsilon(\omega) \). This suggest that most of the properties which apply to \( I_\varepsilon(\omega) \) also apply to \( I_X(\omega) \). Indeed in the following theorem we show that the asymptotic distribution of \( I_X(\omega) \) is exponential with mean and variance \( f(\cdot) \).

By using the above result we now generalise Proposition 8.4.1 to linear processes.

**Theorem 8.4.1** Suppose \( \{X_t\} \) satisfies \( X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_t \), where \( \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \). Let \( I_n(\omega) \) denote the periodogram associated with \( \{X_1, \ldots, X_n\} \) and \( f(\cdot) \) be the spectral density. Then

(i) If \( f(\omega) > 0 \) for all \( \omega \in [0, 2\pi] \) and \( 0 < \omega_1, \ldots, \omega_m < \pi \), then

\[
(I_n(\omega_1)/f(\omega_1), \ldots, I_n(\omega_m)/f(\omega_m))
\]

converges in distribution (as \( n \to \infty \)) to a vector of independent exponential distributions with mean one.

(ii) If in addition \( \mathbb{E}(\varepsilon_t^4) < \infty \) and \( \sum_{j=-\infty}^{\infty} |j^{1/2} \psi_j| < \infty \). Then for \( \omega_j = \frac{2\pi j}{n} \) and \( \omega_k = \frac{2\pi k}{n} \) we have

\[
\text{cov}(I(\omega_k), I(\omega_j)) = \begin{cases} 
2(2\pi)^2 f(\omega_k)^2 + O(n^{-1/2}) & \omega_j = \omega_k = 0 \text{ or } \pi \\
(2\pi)^2 f(\omega_k)^2 + O(n^{-1/2}) & 0 < \omega_j = \omega_k < \pi \\
O(n^{-1}) & \omega_j \neq \omega_k
\end{cases}
\]

where the bound is uniform in \( \omega_j \) and \( \omega_k \).

104
Remark 8.4.1 (Summary of properties of the periodogram for linear processes) \( (i) \)
The periodogram is nonnegative and is an asymptotically an unbiased estimator of the spectral density (when \( \sum_j |\psi_j| < \infty \)).

(ii) Like the spectral density is it symmetric about zero: \( I_n(\omega) = I_n(\omega + \pi) \).

(iii) At the fundamental frequencies \( \{I(\omega_k)\} \) are asymptotically uncorrelated.

(iv) If \( 0 < \omega < \pi \), \( I(\omega) \) is asymptotically exponentially distributed with mean \( f(\omega) \).

We see that the periodogram is extremely erratic and does not converge (in anyway) to the spectral density as \( n \to \infty \). In the following section we discuss this further and consider modifications of the spectral density which lead to a consistent estimate.

8.4.2 Estimating the spectral density

There are several (pretty much equivalent) explanations as to why the raw periodogram is not a good estimator of the spectrum. Intuitively, the simplest explanation is that we have included too many covariance estimators in the estimation of \( f(\omega) \). We see from (8.17) that the periodogram is the Fourier transform of the estimates covariances at \( n \) different lags. Typically the variance for each covariance \( \hat{c}_n(k) \) will be about \( O(n^{-1}) \), hence roughly speaking the variance of \( I_n(\omega) \) will be the sum of these \( n O(n^{-1}) \) variances which leads to a variance of \( O(1) \), which clearly does not converge to zero. This suggest that if we use \( m \) \( (m << n) \) covariances in the estimation of \( f(\omega) \) rather than all \( n \), (where we let \( m \to \infty \)) we may reduce the variance in the estimation (with the cost of introducing some bias) to yield a good estimator of the spectral density. This indeed turns about to be the case.

Another way is to approach the problem is from a nonparametric angle. We note that from Theorem 8.4.1, at the fundamental frequencies \( \{I(\omega_k)\} \) can be treated as uncorrelated random variables, with mean \( f(\omega_k) \) and variance \( f(\omega_k) \). Therefore we can rewrite \( I(\omega_k) \) as

\[
I(\omega_k) = \mathbb{E}(I(\omega_k)) + (I(\omega_k) - \mathbb{E}(I(\omega_k))) \\
\approx f(\omega_k) + f(\omega_k)U_k, \quad k = 1, \ldots, n,
\]

(8.25)

where \( \{U_k\} \) is approximately a mean zero, variance one sequence of uncorrelated random variables and \( \omega_k = 2\pi k/n \). We note that (8.25) resembles the usual nonparametric function plus additive noise, often considered in nonparametric statistics. This suggest that another way to estimate the spectral density us to use a locally weighted average of \( \{I(\omega_k)\} \). Interestingly both the estimation methods mentioned above are practically the same method.

It is worth noting that Parzen (1957) first proposed a consistent method to estimate the spectral density. Furthermore, classical density estimation and spectral density estimation are very similar, and it was spectral density estimation motivated methods, which motivated methods to estimate the density function (one of the first papers on density estimation is Parzen (1962)).

Equation (8.25) motivates the following nonparametric estimator of \( f(\omega) \).

\[
\hat{f}_n(\omega_j) = \sum_k \frac{1}{bn} K\left(\frac{j-k}{bn}\right)I(\omega_k),
\]

(8.26)
where \( W(\cdot) \) is a kernel which satisfies \( \int W(x)dx = 1 \) and \( \int W(x)^2dx < \infty \). An example of \( \hat{f}_n(\omega_j) \) is the local average about frequency \( \omega_j \):

\[
\hat{f}_n(\omega_j) = \frac{1}{bn} \sum_{k=j-bn/2}^{j+bn/2} I(\omega_k).
\]

**Theorem 8.4.2** Suppose \( \{X_t\} \) satisfy \( X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_t \), where \( \sum_{j=-\infty}^{\infty} |j^{1/2} \psi_j| < \infty \) and \( \mathbb{E}(\varepsilon_t^4) < \infty \). Let \( \hat{f}_n(\omega) \) be the spectral estimator defined in (8.26). Then

\[
\mathbb{E}(\hat{f}_n(\omega_j)) \rightarrow f(\omega_j) \quad (8.27)
\]

and

\[
\text{var}(\hat{f}_n(\omega_j)) \rightarrow \begin{cases} 
\frac{1}{2bn} f(\omega_j)^2 & 0 < \omega_j < \pi \\
\frac{1}{bn} f(\omega_j)^2 & \omega_j = 0 \text{ or } \pi.
\end{cases} \quad (8.28)
\]

\( bn \rightarrow \infty, b \rightarrow 0 \) as \( n \rightarrow \infty \).

**Proof.** The proof of both (8.27) and (8.28) are based on the kernel \( K(x/b) \) getting narrow as \( b \rightarrow 0 \), hence there is more localisation as the sample size grows (just like nonparametric regression). We note that since \( f(\omega) = \sigma^2 |\sum_{j=-\infty}^{\infty} \psi \exp(ij\omega)|^2 \), the spectral density, \( f \), is continuous in \( \omega \).

To prove (8.27) we take expectations

\[
|\mathbb{E}(\hat{f}_n(\omega_j)) - f(\omega_j)| = \left| \sum_k \frac{1}{bn} K\left( \frac{k}{bn} \right) \{ \mathbb{E}(I(\omega_j-k)) - f(\omega_j) \} \right|
\]

\[
= \left| \sum_k \frac{1}{bn} K\left( \frac{k}{bn} \right) |\mathbb{E}(I(\omega_j-k)) - f(\omega_j-k)| + \sum_k \frac{1}{bn} K\left( \frac{k}{bn} \right) \{ f(\omega_j) - f(\omega_j-k) \} \right| 
\]

\[
:= I + II.
\]

Using Lemma 8.4.1 we have

\[
I = \sum_k \frac{1}{bn} K\left( \frac{k}{bn} \right) |\mathbb{E}(I(\omega_j-k)) - f(\omega_j-k)| 
\]

\[
\leq K\left( \frac{1}{bn} \sum_k |K\left( \frac{k}{bn} \right)| \right) \times \left( \sum_{|k| \geq n} |c(k)| + \sum_{|k| \leq n} \frac{|k|}{n} |c(k)| \right) \rightarrow 0.
\]

Now we consider

\[
II = \left| \sum_k \frac{1}{bn} K\left( \frac{k}{bn} \right) \{ f(\omega_j) - f(\omega_j-k) \} \right| = O(b).
\]

Since the spectral density \( f(\cdot) \) is continuous, then we have \( II \rightarrow \infty \) as \( bn \rightarrow \infty, b \rightarrow 0 \) and \( n \rightarrow \infty \). The above two bounds mean give (8.27). \( \square \)
We will use Theorem 8.4.1 to prove (8.28). We first assume that \( j \neq 0 \) or \( n \). Evaluating the variance using Theorem 8.4.1 we have

\[
\text{var}(\hat{f}_n(\omega_j)) = \sum_{k_1,k_2} \frac{1}{(bn)^2} K(\frac{j-k_1}{bn})K(\frac{j-k_2}{bn})\text{cov}(I(\omega_{k_1}), I(\omega_{k_2}))
\]

\[
= \sum_k \frac{1}{(bn)^2} K(\frac{j-k}{bn})^2 \text{var}(I(\omega_k)) + O\left(\frac{1}{n}\right)
\]

\[
= \sum_k \frac{1}{(bn)^2} K(\frac{k}{bn})^2 f(\omega_{j-k}) + O\left(\frac{1}{n^{1/2}}\right) \to \frac{1}{bn} f(\omega_j).
\]

A similar proof can be used to prove the case \( j = 0 \) or \( n \). □

The above result means that the mean squared error of the estimator is

\[\mathbb{E}(\hat{f}_n(\omega_j) - f(\omega_j))^2 \to 0\]

as \( bn \to \infty \), \( b \to 0 \) and \( n \to \infty \). Moreover

\[
\mathbb{E}(\hat{f}_n(\omega_j) - f(\omega_j))^2 = O\left(\frac{1}{bn}\right) + \left\{\mathbb{E}(\hat{f}_n(\omega_j)) - f(\omega_j)\right\}^2
\]

\[
= O\left(\frac{1}{bn}\right) + O\left(\sum_{|k| \geq n} |c(k)| + \sum_{|k| \leq n} \frac{|k|}{n} |c(k)| + \int \frac{1}{bn} W(\frac{x}{b}) \cdot (f(\omega) - f(x))dx\right)^2 + O(b).
\]

Hence the rate of convergence depends on the bias of the estimator, in particular the rate of decay of the covariances.

There are several example of kernels that one can use and each has its own optimality property. An interesting discussion on this is given in Priestley (1983), Chapter 6.

As mentioned briefly above we can also estimate the spectrum by truncating the number of covariances estimated. We recall that

\[
I_X(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \hat{c}_n(k) \exp(ik\omega).
\]

Hence a viable estimator of the spectrum is

\[
\hat{f}_n(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \lambda(\frac{k}{M}) \hat{c}_n(k) \exp(ik\omega),
\]

hence \( \lambda(\cdot) \) is a weight function with very little weight (or no weight) for the covariances at large lags. A useful example, using the rectangular function is \( \lambda(x) = 1 \) if \( |x| \leq 1 \) and zero otherwise is

\[
\hat{f}_n(\omega) = \frac{1}{2\pi} \sum_{k=-(M)}^{M} \hat{c}_n(k) \exp(ik\omega),
\]
where $M << n$. Now $\hat{f}_n(\omega)$ has similar properties to $\hat{f}_n(\omega)$. Indeed there is a close relationship between the two which can be seen by using (8.1). Substituting $\hat{c}_n(k) = \frac{1}{2\pi} \int_{0}^{2\pi} I_n(\omega) \exp(-ik\omega) d\omega$ into $\hat{f}_n$ gives

$$
\hat{f}_n(\omega) = \frac{1}{(2\pi)^2} \int I_n(\lambda) \sum_{k=-(n-1)}^{n-1} \lambda(\frac{k}{M}) \exp(ik(\omega - \lambda)) d\lambda = \frac{1}{2\pi} \int I_n(\lambda)K_M(\omega - \lambda) d\omega,
$$

where $K_M(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \lambda(\frac{k}{M}) \exp(ik\omega)$. Now $K_M(\cdot)$ and $\frac{1}{bn}K(\frac{x}{bn})$ (defined in (8.26)) are not necessarily the same function, but they share many of the same characteristics. In other words

$$
K_M(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \lambda(\frac{k}{M}) \exp(ik\omega) = \frac{M}{2\pi} \sum_{k=-(n-1)}^{n-1} \lambda(\frac{k}{M}) \exp(i\frac{k}{M} \cdot M\omega)
$$

$$
\approx MW(M\omega), \quad \text{where } W(\omega) = \int \lambda(x) \exp(i\omega) dx.
$$

Hence $K_M(\omega) = MW(M\omega)$, therefore

$$
\hat{f}_n(\omega) \approx M \int I_n(\lambda)W(M(\omega - \lambda)) d\omega.
$$

Comparing with $\hat{f}_n$ we see that $M$ plays the same role as $b^{-1}$. Furthermore, we observe $\sum_{k} \frac{1}{bn}K(\frac{x}{bn})I(\omega_k)$ is the sum of about $M(\omega_k)$ terms. The equivalent for $K_M(\cdot)$, is that it has the ‘spectral’ width $m = n/M$. In other words since $\hat{f}_n(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \lambda(\frac{k}{M})\hat{c}_n(k) \exp(ik\omega) = \frac{1}{2\pi} \int MI_n(\lambda)W(M(\omega - \lambda)) d\omega$, is the sum of about $m = n/M$ terms.

**Remark 8.4.2 (The distribution of the spectral density estimator)** Using that the periodogram $I_n(\omega)/f(\omega)$ is asymptotically $\chi^2(2)$ distributed and uncorrelated at the fundamental frequencies, we can deduce approximate the distribution of $\hat{f}_n(\omega)$. To obtain the distribution consider the example

$$
\hat{f}_n(\omega_j) = \frac{1}{bn} \sum_{k=j-bn/2}^{j+bn/2} I(\omega_k).
$$

Since $I(\omega_k)/f(\omega_k)$ are approximately $\chi^2(2)$, then since the sum $\sum_{k=j-bn/2}^{j+bn/2} I(\omega_k)$ is taken over a local neighbourhood of $\omega_j$, we have that $f(\omega_j)^{-1} \sum_{k=j-bn/2}^{j+bn/2} I(\omega_k)$ is approximately $\chi^2(2bn)$. Extending this argument to arbitrary kernels we have that $bn f_n(\omega_j)/f(\omega_j)$ is approximately $\chi^2(2bn)$.

We note that when $bn$ is large, then $\chi^2(2bn)$ is close to normal with mean $2bn$ and variance $4bn$. Hence $bn f_n(\omega_j)/f(\omega_j)x$ is approximately normal with mean $2bn$ and variance $4bn$. Therefore

$$
\sqrt{bn} \hat{f}_n(\omega_j) \approx N(2f(\omega_j), 4f(\omega_j)^2).
$$

Using this approximation, we can construct confidence intervals for $f(\omega_j)$. 

108
8.5 The Whittle Likelihood

In Chapter 4 we considered various methods for estimating the parameters of an ARMA process. The most efficient method (when the errors were Gaussian) was the Gaussian maximum likelihood estimator. This estimator was defined in the time domain, but it is interesting to note that a very similar estimator which is asymptotically equivalent to the GMLE estimator can be defined in the frequency domain. We first define the estimator using heuristics to justify it. We then show how it is related to the GMLE (it is the frequency domain approximation of the time domain estimator).

First let us suppose that we observe \( \{X_t\}_{t=1}^{n} \), where

\[
X_t = \sum_{j=1}^{p} \phi_j^{(0)} X_{t-j} + \sum_{j=1}^{q} \theta_j^{(0)} \varepsilon_{t-j} + \varepsilon_t,
\]

and \( \{\varepsilon_t\} \) are iid random variables. As before we will assume that \( \phi_0 = \{\phi_j^{(0)}\} \) and \( \theta_0 = \{\theta_j^{(0)}\} \) are such that the roots of the characteristic polynomial are greater than \( 1 + \delta \). Let us defined the Discrete Fourier Transform

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

We will consider \( \omega \) at the fundamental frequencies \( \omega_k = \frac{2\pi k}{n} \). As we mentioned in Section 8.2 we have

\[
\text{cov}(J_n(\omega_{k_1}), J_n(\omega_{k_2})) = \begin{pmatrix} \text{var}(J_n(\omega_{k_1})) & f(\omega_{k_1}) & k_1 = k_2 \\ f(\omega_{k_1}) & 0 & k_1 \neq k_2 \end{pmatrix}.
\]

Hence if the innovations are Gaussian then \( J_n(\omega) \) is complex Gaussian and we have approximately

\[
J_n(\omega) = \begin{pmatrix} J_n(\omega_1) \\ \vdots \\ J_n(\omega_n) \end{pmatrix} \sim \mathcal{CN}(0, \text{diag}(f(\omega_1), \ldots, f(\omega_n))).
\]

Therefore since \( J_n \) is normally distributed (complex) random vector with mean zero and diagonal matrix variance matrix \( \text{diag}(f(\omega_1), \ldots, f(\omega_n)) \), the log likelihood of \( J_n \) is approximately

\[
L_n^w(\phi, \theta) = \sum_{k=1}^{n} \left( \log |f_{\phi,\theta}(\omega_k)| + \frac{|J_n(\omega_k)|^2}{f_{\phi,\theta}(\omega_k)} \right).
\]

To estimate the parameter we would choose the \( \theta \) and \( \phi \) which minimises the above criterion, that is

\[
(\phi_n^w, \theta_n^w) = \arg \max_{\phi, \theta \in \Omega} L_n^w(\phi, \theta),
\]

where \( \Omega \) consists of all parameters where the roots of the characteristic polynomial have absolute value greater than \( (1 + \delta) \).

Whittle (1962) showed that the above criterion is an approximation of the GMLE. The correct proof is quite complicated and uses several matrix approximations due to Grenander and Szegö (1958). Instead we give a heuristic proof which is quite enlightening.
Remark 8.5.1 (Some properties of circulant matrices)

(i) Let us define the \( n \) dimensional circulant matrix

\[
\Gamma = \begin{pmatrix}
c(0) & c(1) & c(2) & \cdots & c(n-2) & c(n-1) \\
c(n-1) & c(0) & c(2) & \cdots & c(n-3) & c(n-2) \\
c(n-1) & c(1) & c(0) & & & \\
\end{pmatrix}.
\]

See that the elements in each row are the same, just a rotation of each other.

The eigenvalues and vectors of \( \Gamma \) have interesting properties. Define \( f_n(\omega) = \sum_{k=-(n-1)}^{n-1} c(k) \exp(ik\omega) \),

then the eigenvalues are \( \{f_n(\omega_j)\} \) with corresponding eigenvectors \( e_j = (1, \exp(2\pi ij/n), \ldots, \exp(2\pi ij(n-1)/n)) \).

Hence let \( E = (e_1, \ldots, e_n) \), then we can write \( \Gamma = E\Delta E^{-1} \) (8.30)

where \( \Delta = \text{diag}(f(\omega_1), \ldots, f(\omega_n)) \).

(ii) You may wonder about the relevance of circulant matrices to the current setting. However the variance covariance matrix of a stationary process is

\[
\text{var}(X_n) = \Sigma = \begin{pmatrix}
c(0) & c(1) & c(2) & \cdots & c(n-2) & c(n-1) \\
c(1) & c(0) & c(2) & \cdots & c(n-3) & c(n-2) \\
c(n-1) & c(1) & c(0) & & & \\
\end{pmatrix}.
\]

This is a Toeplitz matrix and we observe for large \( n \) it is very close to the circulant matrix, the differences are in the endpoints of the matrix. Hence it can be shown that for large \( n \)

\[
\Sigma \approx E\Delta E^{-1},
\]

and \( E^{-1} \approx \bar{E} \) (where \( \bar{E} \) denotes the complex conjugate of \( E \)).

We will use the results in the lemma above to prove the lemma below. We first observe that in the Gaussian maximum likelihood for the ARMA process can be written as in terms of its spectrum (see 4.10)

\[
L_n(\phi, \theta) = \det |\Sigma(\phi, \theta)| + X_n^T \Sigma(\phi, \theta)^{-1} X_n = \det |\Sigma(f_{\phi, \theta})| + X_n^T \Sigma(f_{\phi, \theta})^{-1} X_n, \quad (8.32)
\]

where \( \Sigma(f_{\phi, \theta})_{s,t} = \int f_{\phi, \theta}(\omega) \exp(i(s-t)\omega) d\omega \) and \( X_n = (X_1, \ldots, X_n) \). We now show that

\[
L_n(\phi, \theta) \approx L_n^w(\phi, \theta).
\]

Lemma 8.5.1 Suppose that \( \{X_t\} \) is a stationary ARMA time series with absolutely summable covariances and \( f_{\phi, \theta}(\omega) \) is the corresponding spectral density function. Then

\[
\det |\Sigma(f_{\phi, \theta})| + X_n^T \Sigma(f_{\phi, \theta})^{-1} X_n \approx \sum_{k=1}^{n} \left( \log |f_{\phi, \theta}(\omega_k)| + \frac{|J_X(\omega_k)|^2}{f_{\phi, \theta}(\omega_k)} \right),
\]

for large \( n \).
PROOF. We give a heuristic proof (details on the precise proof can be found in the remark below). Using (8.31) we have seen that $\Sigma(f_{\phi,\theta})$ can be approximately written in terms of the eigenvalue and eigenvectors of the circulant matrix associated with $\Sigma(f_{\phi,\theta})$, that is

$$\Sigma(f_{\phi,\theta}) \approx E\Delta(f_{\phi,\theta})E$$ \quad and \quad $$\Sigma(f_{\phi,\theta})^{-1} \approx E\Delta(f_{\phi,\theta})^{-1}E,$$

(8.33)

where $\Delta(f_{\phi,\theta}) = \text{diag}(f_n(\omega_1), \ldots, f_n(\omega_n))$, $f^{(n)}_{\phi,\theta}(\omega) = \sum_{j=-(n-1)}^{j=(n-1)} c_{\phi,\theta}(k) \exp(ik\omega) \rightarrow f_{\phi,\theta}(\omega)$ and $\omega_k = 2\pi k/n$. A basic calculations gives that

$$X_nE = (J_n(\omega_1), \ldots, J_n(\omega_n)).$$

(8.34)

Substituting (8.34) and (8.33) into (8.35) yields

$$\frac{1}{n}L_n(\phi, \theta) \approx \frac{1}{n} \sum_{k=1}^{n} \left( \det f^{(n)}_{\phi,\theta}(\omega_k) + \frac{|J_n(\omega_k)|^2}{f_{\phi,\theta}(\omega_k)} \right) = \frac{1}{n} \mathcal{L}^w(\phi, \theta).$$

(8.35)

Hence by using the approximation (8.33) we have derived the Whittle likelihood. This proof was first derived by Tata Subba Rao.

\[ \square \]

Remark 8.5.2 (A rough flavour of the proof) There are various ways to precisely prove this result. All of them show that the Toeplitz matrix can in some sense be approximated by a circulant matrix. This result uses Szegö’s identity (Grenander and Szegö (1958)). Studying the Gaussian likelihood in (8.35) we note that the Whittle likelihood has a similar representation. That is

$$\mathcal{L}^w_n(\phi, \theta) = \sum_{k=1}^{n} \left( \log |f_{\phi,\theta}(\omega_k)| + \frac{|J_n(\omega_k)|^2}{f_{\phi,\theta}(\omega_k)} \right) = \sum_{k=1}^{n} \log |f_{\phi,\theta}(\omega_k)| + X_nU(f_{\phi,\theta}^{-1})X_n,$$

where $U(f_{\phi,\theta})_{s,t} = \int f_{\phi,\theta}(\omega)^{-1} \exp(i(s-t)\omega) d\omega$. Hence we can show that $|U(f_{\phi,\theta}^{-1}) - \Sigma(f_{\phi,\theta})^{-1}| \rightarrow 0$ as $n \rightarrow \infty$ (and its derivatives with respect $\phi$ and $\theta$ also converge), then we can show that $\mathcal{L}^w_n(\phi, \theta)$ and $L_n(\phi, \theta)$ and its derivatives are asymptotically equivalent. Hence the GMLE and the Whittle estimator are asymptotic equivalent. The difficult part in the proof involves showing that $|U(f_{\phi,\theta}^{-1}) - \Sigma(f_{\phi,\theta})^{-1}| \rightarrow 0$ as $n \rightarrow \infty$. It is worth noting that Rainer Dahlhaus has extensively developed this area and considered several interesting generalisations (see for example Dahlhaus (1996), Dahlhaus (1997) and Dahlhaus (2000)).

We now show consistency of the estimator (without showing that its equivalent to the GMLE). To simplify calculations we slightly modify the estimator and exchange the summand in (8.29) with an integral to obtain

$$\mathcal{L}^w(\phi, \theta) = \int_0^{2\pi} \left( \log |f_{\phi,\theta}(\omega_k)| + \frac{|J_n(\omega_k)|^2}{f_{\phi,\theta}(\omega_k)} \right).$$

To estimate the parameter we would choose the $\theta$ and $\phi$ which minimises the above criterion, that is

$$\left( \phi_n^w, \theta_n^w \right) = \arg \max_{\phi,\theta \in \Omega} \mathcal{L}^w_n(\phi, \theta),$$

(8.36)
Lemma 8.5.2 (Consistency) Let us suppose that $d_k(\theta, \phi) = \int f_{\phi, \theta}(\omega) \exp(ik\omega) d\omega$ and $\sum_k |d_k(\theta, \phi)| < \infty$. Let $(\phi_n, \theta_n)$ be defined as in (8.36). Then we have

$$(\phi_n^w, \theta_n^w) \xrightarrow{P} (\theta_0, \phi_0).$$

PROOF. Recall to show consistency we need to show pointwise convergence of $\hat{L}_n$ and equicontinuity. First we show pointwise convergence by evaluating the variance of $\hat{L}_n$, and show that it converges to zero as $n \to \infty$. We first note that by using $d_k(\theta, \phi) = \int f_{\phi, \theta}(\omega) \exp(ik\omega) d\omega$ and using $I_n(\omega) = \frac{1}{2\pi} \sum_{k=-n}^{n-1} \hat{e}_n(k) \exp(ik\omega)$ we can write $L_n^w$ as

$$\frac{1}{n} L_n^w(\phi, \theta) = \int_0^{2\pi} \log|f_{\phi, \theta}(\omega_k)| + \frac{1}{n} \sum_{r=-n}^{n-1} d_r(\theta, \phi) \sum_{k=1}^{n-1} X_k X_{k+r}.$$

Therefore taking the variance gives

$$\text{var}(L_n^w(\phi, \theta)) = \frac{1}{n^2} \sum_{r_1, r_2=-n}^{n-1} d_{r_1}(\theta, \phi) d_{r_2}(\theta, \phi) \sum_{k_1=1}^{n-|r_1|} \sum_{k_2=1}^{n-|r_2|} \text{cov}(X_{k_1}, X_{k_1+r_1}, X_{k_2}, X_{k_2+r_2}) (8.37) \quad \text{We note that}$$

$$\text{cov}(X_{k_1}, X_{k_1+r_1}, X_{k_2}, X_{k_2+r_2}) = \text{cov}(X_{k_1}, X_{k_2}) \text{cov}(X_{k_1+r_1}, X_{k_2+r_2}) + \text{cov}(X_{k_1}, X_{k_2+r_2}) \text{cov}(X_{k_1+r_1}, X_{k_2}) + \text{cov}(X_{k_1}, X_{k_1+r_1}, X_{k_2}, X_{k_2+r_2}).$$

Now we note that $\sum_k |\text{cum}(X_{k_1}, X_{k_1+r_1}, X_{k_2}, X_{k_2+r_2})| < \infty$ and $\sum_k \text{cov}(X_{k_1}, X_{k_2}) < \infty$, hence substituting this and $\sum_k |d_k(\theta, \phi)| < \infty$ into (8.37), we have that

$$\text{var}(L_n^w(\phi, \theta)) = O\left(\frac{1}{n}\right),$$

hence $\text{var}(L_n^w(\phi, \theta)) \to 0$ as $n \to \infty$. Define

$$L_w(\phi, \theta) = \mathbb{E}(L_n^w(\phi, \theta)) = \int \left( \det f_{\phi, \theta}(\omega) + \frac{f_{\phi_0, \theta_0}(\omega)}{f_{\phi_0, \theta}(\omega_k)} \right) d\omega.$$

Hence, since $\text{var}(L_n^w(\phi, \theta)) \to 0$ we have

$$L_n^w(\phi, \theta) \xrightarrow{P} L_w(\phi, \theta).$$

To show equicontinuity we apply the mean value theorem to $L_n^w$. We note that because the parameters $(\phi, \theta) \in \Theta$, have characteristic polynomial whose roots are greater than $(1 + \delta)$ then $f_{\phi, \theta}(\omega)$ is bounded away from zero (indeed there exists a $\delta^* > 0$ where $\inf_{\omega \in \Theta} f_{\phi, \theta}(\omega) \geq \delta^*$). Hence it can be shown that there exists a random sequence $\{K_n\}$ such that $|L_n(\phi_1, \theta_1) - L_n(\phi_2, \theta_2)| \leq K_n(||(\phi_1 - \phi_2), (\theta_1 - \theta_2)||)$ and $K_n$ converges almost surely to a finite constant as $n \to \infty$. Therefore $L_n$ is stochastically equicontinuous (and equicontinuous in probability). Since the parameter space $\Theta$ is compact, all three conditions in Section 5.6 are satisfied and we have consistency of the Whittle estimator.  

We now show asymptotic normality of the Whittle estimator and in the following remark show its relationship to the GMLE estimator.
Let us suppose that $d_k(\theta, \phi) = \int f_\phi(\omega)^{-1} \exp(ik\omega)d\omega$ and $\sum_k |d_k(\theta, \phi)| < \infty$. Let $(\phi_n^w, \theta_n^w)$ be defined as in (8.36)

$$\sqrt{n}(\phi_n^w - \phi_0, \theta_n^w - \theta_0) \overset{D}{\rightarrow} N(0, V^{-1} + VWV^{-1})$$

where

$$V = \frac{1}{2\pi} \int_0^{2\pi} \left( \frac{\nabla f_{\phi,\theta}(\omega)}{f_{\phi,\theta}(\omega)} \right)^2 \frac{\nabla f_{\phi,\theta}(\omega)}{f_{\phi,\theta}(\omega)}' d\omega$$

$$W = \frac{1}{2\pi} \int_0^{2\pi} \left( \frac{\nabla f_{\phi,\theta}(\omega)}{f_{\phi,\theta}(\omega)} \right)^2 \frac{\nabla f_{\phi,\theta}^{-1}(\omega)}{f_{\phi,\theta}^{-1}(\omega)}' f_{\phi,\theta}^{-1}(\omega, -\omega_1, \omega_2),$$

and $f_{\phi,\theta}(\omega_1, \omega_2, \omega_3) = \kappa_4 A(\omega_1)A(\omega_2)A(\omega_3)A(-\omega_1 - \omega_2 - \omega_3)$ is the fourth order spectrum corresponding to the ARMA process with $A(\omega) = \theta_0(\exp(\omega))$. \hfill \Box

**Remark 8.5.3** (i) It is interesting to note that in the case that $\{X_t\}$ comes from a linear time series (such as an ARMA process) then using $f_{\phi,\theta}(\omega, \omega_1, -\omega_2) = \kappa_4 |A(\omega)|^2 |A(\omega_2)|^2 = \frac{\kappa_4}{\kappa_2^2} f(\omega_1)f(\omega_2)$ (for linear processes) we have

$$W = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{2\pi} \left( \frac{\nabla f_{\phi,\theta}(\omega)}{f_{\phi,\theta}(\omega)} \right)^2 \frac{\nabla f_{\phi,\theta}^{-1}(\omega)}{f_{\phi,\theta}^{-1}(\omega)}' \frac{\nabla f_{\phi,\theta}^{-1}(\omega)}{f_{\phi,\theta}^{-1}(\omega)} f_{\phi,\theta}^{-1}(\omega, -\omega_1, \omega_2)$$

$$= \kappa_4 \left( \frac{1}{2\pi} \int_0^{2\pi} \nabla \log f_{\phi,\theta}(\omega) d\omega \right)^2 = \kappa_4 \left( \frac{1}{2\pi} \int_0^{2\pi} \nabla \log (\exp(\omega)) d\omega \right)^2 = 0,$$

where we note that $\int_0^{2\pi} \log f_{\phi,\theta}(\omega) d\omega = 2\pi \log \frac{\sigma_\theta^2}{2\pi}$ by using Kolmogorov’s formula. Hence for linear processes the higher order cumulant plays no role and above theorem reduces to

$$\sqrt{n}(\hat{\phi}_n - \phi_0, \hat{\theta}_n - \theta_0) \overset{D}{\rightarrow} N(0, V^{-1})$$

(ii) Since the GMLE and the Whittle likelihood are asymptotically equivalent they should lead to the same asymptotic distributions. We recall that the GMLE has the asymptotic distribution $\sqrt{n}(\hat{\theta}_n - \theta_0, \hat{\phi}_n - \theta_0) \overset{D}{\rightarrow} N(\Lambda, \sigma_0^2 \Lambda^{-1})$, where $\Lambda = \frac{1}{\sigma_0^2} \begin{pmatrix} \mathbb{E}(U_iU_i') & \mathbb{E}(V_iU_i') \\ \mathbb{E}(U_iV_i') & \mathbb{E}(V_iV_i') \end{pmatrix}$

and $\{U_t\}$ and $\{V_t\}$ are autoregressive processes which satisfy $\phi_0(B)U_t = \varepsilon_t$ and $\theta_0(B)V_t = \varepsilon_t$.

It can be shown that

$$\frac{1}{\sigma_0^2} \int_0^{2\pi} \left( \frac{\nabla f_{\phi,\theta}(\omega)}{f_{\phi,\theta}(\omega)} \right)^2 \frac{\nabla f_{\phi,\theta}^{-1}(\omega)}{f_{\phi,\theta}^{-1}(\omega)}' d\omega.$$