4.1 Power Series and Euler's Formula

Taylor series (or, in general, *power series*) are the quintessential representations. For example, it is possible to represent rather arbitrary functions as infinite polynomials. This is useful for differentiating and integrating, as well as finding limits, and deriving other formulæ.

Theorem 4.1.1 (Taylor's Theorem). *Suppose that the function f is* analytic *in the disk $\mathbb{D}_R := \{z \in \mathbb{C} : |z - z_0| < R\}$. Then f has a Taylor series representation*

$$f(z_0) + f'(z_0)(z - z_0) + \frac{f''(z_0)}{2!}(z - z_0)^2 + \cdots = \sum_{j=0}^{\infty} \frac{f^{(j)}(z_0)}{j!}(z - z_0)^j$$

which converges to $f(z)$ for all $z \in \mathbb{D}_R$. Furthermore, if $R' < R$, then the convergence is uniform for any z in the closed disk $\overline{D}_{R'} = \{|z - z_0| \leq R'\}$.

There is a partial converse to Taylor's Theorem which says that a convergent power series *defines* an analytic function.

Theorem 4.1.2 (Power Series). *Suppose that the power series*

$$\sum_{j=0}^{\infty} a_j z^j$$

has radius of convergence $R \neq 0$. Then the function f defined by

$$f(z) := \sum_{j=0}^{\infty} a_j z^j$$

is analytic inside the disk $\{|z| < R\}$.

For instance, recall that

- $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$
- $\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$
- $\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}$

Example 4.1.3. Suppose that $i = \sqrt{-1}$ is the imaginary unit. Then,

$$\begin{aligned} e^{ix} = \sum_{k=0}^{\infty} \frac{(ix)^k}{k!} &= \sum_{k=0}^{\infty} i^{2k} \frac{x^{2k}}{(2k)!} + \sum_{k=0}^{\infty} i^{2k+1} \frac{x^{2k+1}}{(2k+1)!} \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} + \sum_{k=0}^{\infty} i(-1)^k \frac{x^{2k+1}}{(2k+1)!} \\ &= \cos x + i \sin x \end{aligned}$$

which is the celebrated *Euler formula*.

In particular, if $x = \pi$, then

$$e^{i\pi} = \cos\pi + i\sin\pi = -1$$

or

$$e^{i\pi} + 1 = 0$$

which relates the FIVE most important numbers in one simple formula!!!!

Exercise 4.1.4. Fill in any missing steps for yourself in the derivation of the *Euler formula*, especially the step from the second to third equality.

4.2 Fourier Analysis

Another useful represention theorem comes from Fourier analysis which says that particular periodic functions can be expressed as sums of sines and cosines. The subject of *wavelets* endeavours to extend these Fourier representations.

We will definitely be discussing representations in terms of sines and cosines (also called *harmonic representations*).

Theorem 4.2.1 (Fourier Convergence Theorem). *Suppose that f and f' are piecewise continuous on the interval $-L \leq x < L$. Further, suppose that f is defined outside the interval $[-L, L)$ so that it is periodic with period $2L$. Then f has a Fourier series representation*

$$g(x) = \frac{a_0}{2} + \sum_{m=1}^{\infty}\left(a_m \cos\frac{m\pi x}{L} + b_m \sin\frac{m\pi x}{L}\right), \tag{†}$$

whose coefficients are given by

$$a_m = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{m\pi x}{L}dx, \qquad m = 0, 1, 2, \dots,$$

and

$$b_m = \frac{1}{L}\int_{-L}^{L} f(x)\sin\frac{m\pi x}{L}dx, \qquad m = 1, 2, \dots.$$

The Fourier series $g(x)$ given in (†) converges to $f(x)$ at all points where $f(x)$ is continuous, and to $[f(x+) + f(x-)]/2$ at all points where f is discontinuous.

This really is a very powerful theorem. For reinforcement, you should work through the following two exercises. The mathematics is really nothing more than integration-by-parts from elementary calculus; the only "hard" part is keeping track of the indices.

Exercise 4.2.2. Suppose that

$$f_0(x) = \begin{cases} 1, & 0 \leq x < 2, \\ -1, & -2 \leq x < 0. \end{cases}$$

Let f be the periodic extension of f_0 to all of $\mathbb{R}$. Verify that f satisfies the hypotheses of the Fourier Convergence Theorem, and then compute its Fourier Series $g(x)$. State the limiting values of $g(x)$ for all $x \in \mathbb{R}$.

Exercise 4.2.3. Suppose that

$$f_0(x) = |x|, \quad -\pi \le x < \pi.$$

Let f be the periodic extension of f_0 to all of $\mathbb{R}$. Verify that f satisfies the hypotheses of the Fourier Convergence Theorem, and then compute its Fourier Series $g(x)$. State the limiting values of $g(x)$ for all $x \in \mathbb{R}$.

4.3 Decomposing Stochastic Processes

It is sometimes useful to write an arbitrary random variable X in the form $X = A + M$ where A and M are themselves random variables, except that they possess certain "nice" properties.

Example 4.3.1. A famous example from martingale theory is the Doob Decomposition Theorem which states that if the stochastic process $\{X_t\}$ is a *submartingale*, then it can be written in the form

$$X_t = X_0 + A_t + M_t$$

where M_t is a *(local) martingale*, A_t is a *previsible increasing process*, and $M_0 = A_0 = 0$. This deep result provides the foundation for the theory of *stochastic integration*.

However, we will only be interested in decomposing stationary stochastic processes $\{X_t\}$. In this case, our decompositions will be of the form

$$X_t = S(t) + R_t$$

where $S(t)$ will be a "smooth" non-random function and R_t will be a "rough" stochastic process. However, the rough process R will be sufficiently "nice" that it can be easily analyzed (such as "White Noise"). Note that we sometimes write the function as $S(t)$, and sometimes as S_t.

In Section 2.6 of [2] we will learn about the *Wold Decomposition* of a *nondeterministic* stationary process.

4.4 Classical Decomposition

Suppose that $\{X_t\}$ is a stochastic process defined by the *classical decompostion*. That is, X is given by

$$X_t = m_t + s_t + Y_t$$

where $m_t = m(t)$ is a slowly changing function called the *trend component*, $s_t = s(t)$ is a function with known period d called the *seasonal component*, and Y_t is a weakly stationary stochastic process called the *random noise component*. In accordance with our earlier section on decompositions, $m_t + s_t$ is "smooth" and Y_t is "rough."

Recall that we defined a time series model in Definition 2.2.1. Two important models which will serve us well are the following.

Non-seasonal Model with Trend

$$X_t = m_t + Y_t$$

where $\mathbb{E}(Y_t) = 0$ for all t.

Classical Decompostion Model

$$X_t = m_t + s_t + Y_t$$

where $\mathbb{E}(Y_t) = 0$ for all t, $s_{t+d} = s_t$, and $\sum_{j=1}^{d} s_j = 0$.

Note that $s_{t+d} = s_t$ says that the function s is periodic with period d. (Think back to the Fourier Convergence Theorem.)

5 More on Stationary Processes

5.1 Strongly Stationary Stochastic Processes

To begin this section, recall that the concept of a weakly stationary stochastic process was introduced in Definition 3.3.1 of Section 3.3.

Definition 3.3.1. We call the stochastic process $\{X_t\}$ *second-order (or weakly) stationary* if

- there is a constant μ such that $\mu(t) = \mu$ for all t, and
- $\gamma(t+h, t)$ only depends on h; that is, if $\gamma(t+h,t) = \gamma(h)$ for all t and for all h.

As you might have guessed, there is such a thing as a *strongly* stationary process.

Definition 5.1.1. A stochastic process $\{X_t\}$ is *strongly stationary* if for any $m \in \mathbb{N}$ and for any times $t_1, \ldots, t_m$, the joint distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_m})$ is the same as the joint distribution of $(X_{t_1+s}, X_{t_2+s}, \ldots, X_{t_m+s})$ for any s.

Remark. As noted in the textbook [2], if $\{X_t\}$ is a strongly stationary process, then it must necessarily be weakly stationary. You will be asked to prove this in Problem 1.3.

Remark. The condition of strong stationarity is actually quite strict, and is of little use in practice; weak stationarity is much more useful. Hence, unless otherwise noted, whenever we refer to a process as *stationary* without qualification, it will be by default a weakly stationary process.

Here are some important properties of strongly stationary processes.

Fact. Suppose that $\{X_t\}$ is a strongly stationary process.

- The random variables X_t are identically distributed.
- The distribution of the random vector (X_t, X_{t+h}) is the same as the distribution of the random vector (X_1, X_{1+h}) for all t and h.
- If $X_t \in L^2$ for each t, then $\{X_t\}$ is weakly stationary.
- There exist processes which are weakly stationary, but not strongly stationary. (i.e., weak stationarity does not imply strong stationarity)
- An iid sequence is strictly stationary.

Exercise 5.1.2. You should try to prove these five properties of strongly stationary processes. If you get stuck, consult pages 49–50 of [2].

Although it has not yet been made clear, the concept of stationarity is extremely important for the analysis of time series. However, it must be noted that since we are defining a time series to be one realization of a (possibly imaginary) stochastic process, the question of stationarity is really a very subtle one. You can actually never tell if a time series is stationary *after* you have observed it. The real question is:

Should we MODEL the time series as a realization of a stationary process?

5.2 The Autoregressive Moving Average ($ARMA$) Processes

Moving average (MA) processes, autoregressive (AR) processes, and autoregressive moving average ($ARMA$) processes provide three interesting classes of stationary processes. Although they are usually of limited value by themselves in practice when the goal is to develop an understanding of the underpinings of a particular model, they are of much more value when the goal of modelling is no more than a means to an end, such as in *forecasting.* After we learn about spectral analysis, we will see that the second order properties of a stationary time series can be well-approximated by that of an appropriate $ARMA$ process. For now, however, we will study them for their own sake!

Example 5.2.1 ($MA(1)$, First-order Moving Average Process). Suppose that $\{X_t\}$ is a discrete time stochastic process defined recursively by

$$X_t = Z_t + \theta Z_{t-1}, \quad t = 0, 1, \ldots$$

where $\theta \in \mathbb{R}$ and $\{Z_t\}$ is white noise with variance σ^2. Clearly $\mathbb{E}(X_t) = 0$ for all t, and $\text{Var}(X_t) = \mathbb{E}(X_t^2) = \sigma^2(1+\theta^2) < \infty$. Thus we find that

$$\gamma(t+h, t) = \begin{cases} \sigma^2(1+\theta^2), & h = 0, \\ \sigma^2\theta, & h = \pm 1, \\ 0, & |h| > 1. \end{cases}$$

As a result, we conclude that the $MA(1)$ process $\{X_t\}$ is stationary. Since $\gamma(0) = \sigma^2(1+\theta^2)$, we see that

$$\rho(h) = \begin{cases} 1, & h = 0, \\ \theta/(1+\theta^2), & h = \pm 1, \\ 0, & |h| > 1. \end{cases}$$

Example 5.2.2 ($AR(1)$, First-order Autoregressive Process). Suppose that $\{X_t\}$ is a discrete time stochastic process defined recursively by

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, 1, \ldots, \tag{$\dagger$}$$

where $\phi \in \mathbb{R}$ and $\{Z_t\}$ is white noise with variance σ^2. We shall show later that $\{X_t\}$ is stationary if and only if $|\phi| < 1$ (i.e., if $-1 < \phi < 1$).

Assuming $|\phi| < 1$, we now find the ACF of the $AR(1)$ process. By stationarity, $\mathbb{E}(X_t) = \mu$ does not depend on t. Since $\mathbb{E}(Z_t) = 0$, we find that taking expectations gives

$$\mathbb{E}(X_t) = \phi\mathbb{E}(X_{t-1}) + \mathbb{E}(Z_t) \quad \text{or} \quad \mu = \phi\mu,$$

which is only possible if $\mu = 0$.

Next, if we multiply both sides of (†) by X_{t-k}, then taking expectations and dividing by $\mathrm{Var}(X_t) = \gamma(0)$, we find that

$$\rho(k) = \phi\rho(k-1).$$

Since $\rho(0) = 1$, we find that this difference equation implies that

$$\rho(k) = \phi^k, \quad k = 0, 1, \ldots.$$

Remark. You should see a similarity between the $AR(1)$ in (†) and the random walk of Example 3.4.2. That is, if $\{Z_t\}$ is not white noise but rather iid noise, and $\phi = 1$, then (†) is a random walk. In this case, it is *not* true that $|\phi| < 1$, so we conclude that the random walk is *not* stationary. This is no surprise since we calculated directly that $\gamma(t+h, t) = t\sigma^2$ in Example 3.4.2.

Recall that the ACF is a measure of the serial dependence within a stochastic process. As we saw in Examples 3.3.2 and 3.3.3, two stationary processes may share the same second order properties. The following example provides a more substantial illustration of this fact.

Example 5.2.3 (Limitations of ACF for summarizing serial dependence). Consider the $AR(1)$ process of (†), except that instead of $\{Z_t\}$ being white noise, suppose that Z_t is defined as the two component mixture

$$Z_t = \begin{cases} 0, & \text{with probability } p, \\ U_t, & \text{with probability } 1-p. \end{cases} \tag{‡}$$

where $\{U_t\}$ is a normally distributed iid noise sequence. Notice that if p is near 1, then realizations of this new process differ from the original $AR(1)$ process by their inclusion of exponentially decaying subsequences which correspond to successive $Z_t = 0$. However, this leaves the autocorrelation function of $\{X_t\}$ unchanged!

Exercise 5.2.4. Prove that the ACF of the process defined in (‡) is $\rho(k) = \phi^k$, for $k = 0, 1, \ldots$, the same as an $AR(1)$ process.

Example 5.2.5 ($MA(q)$, q-th order Moving Average Process). Suppose that $\{X_t\}$ is a discrete time stochastic process defined recursively by

$$X_t = Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j}, \quad t = 0, 1, \ldots$$

where $\theta_j \in \mathbb{R}$ and $\{Z_t\} \sim WN(0, \sigma^2)$. We call $\{X_t\}$ a *moving average process of order q*.

Example 5.2.6 ($AR(p)$, p-th order Autoregressive Process). Suppose that $\{X_t\}$ is a discrete time stochastic process defined recursively by

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t, \quad t = 0, 1, \ldots, \tag{$*$}$$

where $\phi_k \in \mathbb{R}$ and $\{Z_t\} \sim WN(0, \sigma^2)$. We call $\{X_t\}$ an *autoregressive process of order p*. Notice that in the $AR(q)$ process ($*$), each X_t is defined in terms of its predecessors X_s, $s < t$.

Example 5.2.7 (*ARMA*(p, q)**, Autoregressive Moving Average Process).** If we combine the $MA(q)$ and $AR(p)$ processes, then we obtain the so-called *autoregressive moving average process.* Suppose that $\{X_t\}$ is a discrete time stochastic process defined by

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j}, \quad t = 0, 1, \ldots.$$

where θ_j, $\phi_k \in \mathbb{R}$ and $\{Z_t\} \sim WN(0, \sigma^2)$. Then $\{X_t\}$ is called an autoregressive moving average process, which is often abbreviated $\{X_t\} \sim ARMA(p, q)$. Notice that if $p = 0$ or $q = 0$, then the definition of the $ARMA(p, q)$ process reduces to $MA(q)$ or $AR(p)$, respectively. An equivalent way to write the $ARMA(p, q)$ process $\{X_t\}$ is as follows:

$$X_t - \sum_{k=1}^{p} \phi_k X_{t-k} = Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j}.$$

The stationarity of an $ARMA(p, q)$ process is rather subtle, as was indicated in Example 5.2.2. Since the moving average processes are always stationary, the question of stationarity really just depends on the autoregressive part.

Theorem 5.2.8 (Stationarity of the $AR(p)$ **process).** *The process* $\{X_t\} \sim AR(p)$ *defined by*

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t, \quad t = 0, 1, \ldots,$$

where $\{Z_t\} \sim WN(0, \sigma^2)$, *is stationary if and only if the* complex *polynomial*

$$\phi(z) := 1 - \sum_{k=1}^{p} \phi_k z^k = 1 - \phi z - \phi_2 z^2 - \cdots - \phi_k z^k$$

has no roots for all $z \in \mathbb{C}$ *with* $|z| = 1$. *That is,* $\{X_t\}$ *is stationary if* $\phi(z) \neq 0$ *for every* $|z| = 1$.

Remark. The necessity of complex numbers arises from the fact that any polynomial of degree k has k complex roots (counting multiplicity). This is known as the *Fundamental Theorem of Algebra.* It is easy to see that limiting polynomial roots to just the real numbers is inadequate. For example, the polynomial $\phi(z) = 1 + z^2$ has no real roots. However, it *does* have two complex roots, namely $z = \sqrt{-1} = i$ and $z = -\sqrt{-1} = -i$. The polynomial $\phi(z) = 1 - z^2$ also has *two* complex roots, namely $z = 1$ and $z = 1$. That is, counting multiplicity, there are two roots.

We can now extend the above theorem to $ARMA(p, q)$ processes.

Theorem 5.2.9 (Stationarity of the $ARMA(p, q)$ **process).** *The process* $\{X_t\} \sim ARMA(p, q)$ *defined by*

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j}, \quad t = 0, 1, \ldots,$$

where $\{Z_t\} \sim WN(0, \sigma^2)$, *is stationary if and only if the* complex *polynomial*

$$\phi(z) := 1 - \sum_{k=1}^{p} \phi_k z^k = 1 - \phi z - \phi_2 z^2 - \cdots - \phi_k z^k$$

has no roots for all $z \in \mathbb{C}$ with $|z| = 1$. That is, $\{X_t\}$ is stationary if $\phi(z) \neq 0$ for every $|z| = 1$. This is equivalent to the condition that the complex polynomials $\phi(z)$ and

$$\theta(z) := 1 - \sum_{j=1}^{q} \theta_j z^j = 1 - \theta z - \theta_2 z^2 - \cdots - \theta_k z^k$$

have no common factors.

6 Introduction to the Spectral Analysis of Stationary Processes

6.1 A Quick Review

In high school, when you learned to graph trignometric functions you were probably shown what to do with expressions of the form

$$a\cos(kt + b).$$

You will certainly recall that a is the *amplitude*, b/k is the *phase shift*, and k is the *frequency* (so that $p = 2\pi/k$ is the *period*).

Exercise 6.1.1. Sketch a graph of the function $f(t) = -2\cos(2t + \pi/4)$.

While this form is excellent for interpretation (and graphing), another form proves useful for time series analysis. Motivated by the form of the Fourier series, we see that an application of a simple trig formula gives

$$\begin{aligned} a\cos(kt + b) &= a(\cos b\cos(kt) - \sin b\sin(kt)) \\ &= a\cos b\cos(kt) - a\sin b\sin(kt) \\ &= \alpha\cos(kt) + \beta\sin(kt) \end{aligned}$$

where we have written $\alpha := a\cos b$ and $\beta := -a\sin b$.

Notice also that $\alpha^2 + \beta^2 = a^2\cos^2 b + a^2\sin^2 b = a^2$ and

$$-\frac{\beta}{\alpha} = \frac{a\sin b}{a\cos b} = \tan b.$$

Thus, the inverse relationship of α and β to a and b is

$$a = \sqrt{\alpha^2 + \beta^2} \quad \text{and} \quad b = \arctan\left(-\frac{\beta}{\alpha}\right).$$

Remark. You will notice a suspicious similarity to the polar coordinates change-of-variables.

Exercise 6.1.2. Sketch a graph of the function $g(t) = -\sqrt{2}\cos(2t) + \sqrt{2}\sin(2t)$.

Exercise 6.1.3. With f and g as in Exercise 6.1.1 and Exercise 6.1.2, respectively, compute both $\int f(t)\,dt$ and $\int g(t)\,dt$. What can you say about these two answers? Why?

6.2 Fourier Frequencies

It turns out that it will be necessary to restrict our possible frequencies to the so-called *Fourier frequencies.* In practical applications, we shall need to consider only frequencies in the range $0 \le k \le \pi$. If $k = \pi$, then this corresponds to an alternating sequence of positive and negative values. We further restrict k to be of the form

$$k = \frac{2\pi j}{n} \quad \text{for some positive integer} \quad j < \frac{n}{2}. \tag{$\dagger$}$$

Definition 6.2.1. The discrete set of frequencies defined by (†) is called the *Fourier frequencies.*

We now state an extremely useful result which is one of the reasons for considering the Fourier frequencies.

Theorem 6.2.2. *If $n \in \mathbb{N}$ and $k = 2\pi j/n$ for any positive integer $j < n/2$, then*

- $\displaystyle\sum_{t=1}^{n} \cos(kt) = \sum_{t=1}^{n} \sin(kt) = \sum_{t=1}^{n} \cos(kt)\sin(kt) = 0;$
- $\displaystyle\sum_{t=1}^{n} \cos^2(kt) = \sum_{t=1}^{n} \sin^2(kt) = \frac{n}{2}.$

Exercise 6.2.3. Use the facts that $e^{ix} = \cos x + i \sin x$, and

$$\sum_{t=1}^{n} z^t = z \cdot \frac{1 - z^n}{1 - z}$$

for any $z \neq 1$, as well as the trigonometric identities $2\cos^2 x = 1 + \cos(2x)$ and $2\sin^2 x = 1 - \cos(2x)$ to prove the above theorem. (You might find it helpful to recall that the complex number $a + bi = 0$ if and only if both $a = 0$ and $b = 0$.)

Notice that for the Fourier frequencies, since $j < n/2$, we have $k \le \pi$. Thus, the possible values for the frequency k lie in the range

$$\frac{2\pi}{n} \le k \le \pi,$$

and therefore the possible values for the period $p = 2\pi/k$ lie in the range $2 \le p \le n$.

We remark that if $p = 1$, then there is no cyclic phenomenon, and if $p > n$, then we cannot tell if there is cyclic behaviour or if we are just observing a trend.

Finally, note that j is the number of complete cycles observed, which is one reason that it is convenient to assume an integer number of cycles.

Exercise 6.2.4. If you are not comfortable with this manipulation of sinusoidal functions, then sketch graphs illustrating the different scenarios listed above.

6.3 Some Multiple Linear Regression

Recall our basic decomposition: $y(t) = s(t) + z(t)$ where s is a smooth signal and z is noise. Using the assumption that the signal can be expressed as a sinusoidal wave, we find

$$y(t) = s(t) + z(t) = a\cos(kt + b) + z(t) = \alpha\cos(kt) + \beta\sin(kt) + z(t).$$

If we evaluate this equation at times $t_1, t_2, \ldots, t_n$ we find that

$$\begin{aligned} y(t_1) &= \alpha\cos(kt_1) + \beta\sin(kt_1) + z(t_1) \\ y(t_2) &= \alpha\cos(kt_2) + \beta\sin(kt_2) + z(t_2) \\ &\vdots \\ y(t_n) &= \alpha\cos(kt_n) + \beta\sin(kt_n) + z(t_n) \end{aligned}$$

which we can rewrite as

$$\begin{aligned} y_1 &= \alpha x_{11} + \beta x_{12} + \epsilon_1 \\ y_2 &= \alpha x_{21} + \beta x_{22} + \epsilon_2 \\ &\vdots \\ y_n &= \alpha x_{n1} + \beta x_{n2} + \epsilon_n. \end{aligned}$$

Notice that this looks suspiciously like the linear regression model

$$\boldsymbol{Y} = \boldsymbol{X}\begin{pmatrix}\alpha \\ \beta\end{pmatrix} + \boldsymbol{\epsilon}.$$

where

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{pmatrix} = \begin{pmatrix} \cos(kt_1) & \sin(kt_1) \\ \cos(kt_2) & \sin(kt_2) \\ \vdots & \vdots \\ \cos(kt_n) & \sin(kt_n) \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

(See Sections 11.10 and 11.11 in [5].)

One consequence of these computations combined with Theorem 6.2.2 is that there is no collinearity among the x-variables.

Theorem 6.3.1. *If*

$$\boldsymbol{X}_1 = \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} \quad \textit{and} \quad \boldsymbol{X}_2 = \begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{pmatrix},$$

then $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are statistically independent. That is, $\boldsymbol{X}_1 \perp \boldsymbol{X}_2$. Furthermore, $\boldsymbol{X}_1 \perp \mathbf{1}$ and $\boldsymbol{X}_2 \perp \mathbf{1}$, where $\mathbf{1}$ is the column vector $\mathbf{1} = (1\ 1\ \cdots\ 1)^T$ (n times).

Exercise 6.3.2. Use Theorem 6.2.2 to prove this theorem. Recall that vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal, written $\boldsymbol{u} \perp \boldsymbol{v}$, if their dot product is zero, i.e., $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

Suppose now that there is no noise so that $\boldsymbol{\epsilon} = 0$. As a result of the above theorem, inferences about α and β are independent of each other. (They are orthogonal.) We also know from the theory of linear regression that the least squares estimator of α, β is given by

$$\begin{pmatrix}\hat{\alpha} \\ \hat{\beta}\end{pmatrix} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}.$$

Now,

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{pmatrix} \sum_{t=1}^{n} \cos^2(kt) & \sum_{t=1}^{n} \cos(kt)\sin(kt) \\ \sum_{t=1}^{n} \sin(kt)\cos(kt) & \sum_{t=1}^{n} \sin^2(kt) \end{pmatrix} = \begin{pmatrix} n/2 & 0 \\ 0 & n/2 \end{pmatrix}.$$

Thus we find that

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \begin{pmatrix} 2/n & 0 \\ 0 & 2/n \end{pmatrix} \boldsymbol{X}^T\boldsymbol{Y} = \begin{pmatrix} \frac{2}{n}\sum_{t=1}^{n} y_t \cos(kt) \\ \frac{2}{n}\sum_{t=1}^{n} y_t \sin(kt) \end{pmatrix}$$

which tells us that

$$\hat{\alpha} = \frac{2}{n}\sum_{t=1}^{n} y_t \cos(kt) \quad \text{and} \quad \hat{\beta} = \frac{2}{n}\sum_{t=1}^{n} y_t \sin(kt).$$

Recall that the *regression sum of squares* is given by $\boldsymbol{Y}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$. From this we find that

$$\boldsymbol{Y}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \frac{2}{n}\left[\left(\sum_{t=1}^{n} y_t \cos(kt)\right)^2 + \left(\sum_{t=1}^{n} y_t \sin(kt)\right)^2\right] = \frac{n}{2}\left(\hat{\alpha}^2 + \hat{\beta}^2\right),$$

and we say that the regression sum of squares is on *two degrees of freedom* (one for α and one for β).

6.4 The Periodogram

The purpose of the previous section was to motivate the following definition.

Definition 6.4.1. If we define

$$I(k) := \frac{1}{n}\left[\left(\sum_{t=1}^{n} y_t \cos(kt)\right)^2 + \left(\sum_{t=1}^{n} y_t \sin(kt)\right)^2\right] = \frac{1}{n}\left|\sum_{t=1}^{n} y_t e^{-itk}\right|^2 \qquad (\dagger)$$

for $0 \le k \le \pi$, then the plot of $I(k)$ vs. k is called the *periodogram* of the time series $\{y_t\}$.

Comparing the definition of the periodogram I to the regression sum of squares, we find that

$$\begin{aligned} I(k) &= \frac{1}{n}\left(\left(\sum y_t \cos(kt)\right)^2 + \left(\sum y_t \sin(kt)\right)^2\right) \\ &= \frac{1}{2}\left(\text{regression sum of squares for } \alpha, \beta\right). \end{aligned}$$

In other words, we have the correspondence

$$\text{Sum of Squares} \equiv \text{Amplitude} \equiv \text{Periodogram}.$$

Remark. Although the definition of $I(k)$ given in (†) makes sense *for any* k, if the period k is not a Fourier frequency, then I no longer represents the regression sum of squares.

Exercise 6.4.2. Suppose that $\{y_t\}$ is a time series with periodogram I. Show that $I(0) = n\bar{y}^2$.

Notice that we can also decompose the *total variation* in the time series $\{y_t\}$ as

$$\sum_{t=1}^{n} y_t^2 = I(0) + 2\sum_{j=1}^{m} I(2\pi j/n) + I(\pi)$$

where m is the largest integer less than $n/2$, and the term $I(\pi) = 0$ is n is odd.

Remark. The periodogram is useful for exploring cyclic patterns in time series data. This is the simple reason for excluding the period $k = 0$ in the definition of I.

6.5 The Spectrum

We are now prepared to delve further into the Fourier analysis of a stationary time series.

Definition 6.5.1. Suppose that $\{X_t\}$ is a discrete time stationary process with autocovariance function $\gamma_h := \gamma(h)$. The *autocovariance generating function* is the function $G : \mathbb{C} \to \mathbb{C}$ given by

$$G(z) := \sum_{h=-\infty}^{\infty} \gamma_h z^h.$$

Definition 6.5.2. The *spectrum (or spectral density)* of the discrete time stationary process $\{X_t\}$ is the function $f : \mathbb{R} \to \mathbb{C}$ defined by $f(\lambda) := G(e^{-i\lambda})$. In other words, if we consider the ACVF generating function and choose $z = e^{-i\lambda}$ with $\lambda \in \mathbb{R}$, then the spectrum of $\{X_t\}$ is

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_h e^{-ih\lambda}. \tag{†}$$

Remark. There is a technical point to note about the definition of the spectrum, namely the series defined by (†) may not converge. If we restrict ourselves to those stationary processes with $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then the series in (†) converges absolutely.

If you are familiar with Fourier analysis, then the spectrum should remind you of the *discrete Fourier transform*. It might also remind you of a *characteristic function* from your earlier probability courses.

Since $\gamma(h) = \gamma(-h)$, and since $e^{i\lambda} + e^{-i\lambda} = 2\cos\lambda$, we can also express the spectrum as

$$f(\lambda) = \gamma_0 + 2\sum_{h=1}^{\infty} \gamma_h \cos(h\lambda).$$

Definition 6.5.3. If $\gamma_0 = \text{Var}(X_t) = \sigma^2$, then the normalized spectrum is defined to be

$$f^*(\lambda) := \frac{f(\lambda)}{\sigma^2} = \sum_{h=-\infty}^{\infty} \rho_h e^{-ih\lambda} = 1 + 2\sum_{h=1}^{\infty} \rho_h \cos(h\lambda)$$

where $\rho_h := \rho(h)$ is the ACF of $\{X_t\}$.

Exercise 6.5.4. Show that if $\{X_t\} \sim WN(0,1)$ is white noise, then $f^*(\lambda) = 1$ for all λ. The fact that the spectrum of white noise is constant is analogous to the flat optical spectrum of white light, and is the source of the luminous name *White Noise*.

Recall that the periodogram of the observed time series $\{x_t\}$ is given by

$$I(\lambda) = \hat{\gamma}_0 + 2\sum_{h=1}^{n-1} \hat{\gamma}_h \cos(h\lambda)$$

where $\hat{\gamma}_h$ is used to estimate γ_h. This shows that the periodogram is may be viewed as an estimator of the spectrum.

Fact. From the definition of $f(\lambda)$ we see that

- $f(\lambda) = f(-\lambda)$,
- $f(\lambda) = f(\lambda + 2\pi m)$ for all integers m, and
- $f(\lambda) \geq 0$ for all $\lambda \in [-\pi, \pi)$.

Together these imply that it is necessary to define $f(\lambda)$ only for $\lambda \in [0, \pi]$. Recall further that we also restricted the periodogram to frequencies $\lambda \in [0, \pi]$.

As already noted in Definition 6.5.2, f is the discrete Fourier transform of γ. The Fourier Inversion Theorem tells us that we can recover γ from f. Hence,

$$\gamma(h) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda)\, d\lambda = \frac{1}{\pi}\int_{0}^{\pi} \cos(h\lambda) f(\lambda)\, d\lambda. \tag{$\dagger$}$$

In fact, this equation characterizes the class of legitimate ACVFs.

Remark. Wold's Theorem states that $\gamma(h)$ is a legitimate ACVF if and only if there exists a non-negative function $f(\lambda)$ on $(0, \pi)$ such that γ has the representation ($\dagger$).

Remark. Anytime you ask someone to *define* the Fourier transform, there is always a question of what to do with the "2π"s. We have (arbitrarily) decided that they will appear with the inverse formula.

As a final note, we observe that, mathematically speaking, both the spectrum and the ACVF contain *exactly* the same information about the stationary process in question. However, from a pragmatic point of view, the spectrum can be interpreted in terms of the inherent tendency for realizations of the process to exhibit cyclic variations about the mean.

Remark. Any (non-negative, integrable) function on $(0, \pi)$ can be a spectrum, but *not* every function on the integers can be an ACVF.

We end this section by computing some normalized spectra.

Example 6.5.5 ($f^*(\lambda)$ for $MA(1)$). Recall from Example 5.2.1 that the autocorrelation function of the $AR(1)$ process is

$$\rho_h = \rho(h) = \begin{cases} 1, & h = 0, \\ \theta/(1+\theta^2), & h = \pm 1, \\ 0, & |h| > 1. \end{cases}$$

Thus we find that the normalized spectrum is

$$f^*(\lambda) = \sum_{h=-\infty}^{\infty} \rho_h e^{-ih\lambda} = \frac{\theta}{1+\theta^2} e^{i\lambda} + 1 + \frac{\theta}{1+\theta^2} e^{-i\lambda} = 1 + \frac{2\theta}{1+\theta^2}\cos(\lambda) = \frac{1+\theta^2+2\theta\cos(\lambda)}{1+\theta^2}$$

Example 6.5.6 ($f^*(\lambda)$ for $AR(1)$). Recall from Example 5.2.2 that the autocorrelation function of the $AR(1)$ process is $\rho(h) = \phi^h$, $h = 0, 1, \ldots$. Thus we find that the normalized spectrum is

$$\begin{aligned}
f^*(\lambda) = \sum_{h=-\infty}^{\infty} \phi^h e^{-ih\lambda} = \sum_{h=-\infty}^{\infty} \left(\phi e^{-i\lambda}\right)^h &= \sum_{h=-\infty}^{-1} \left(\phi e^{-i\lambda}\right)^h + \sum_{h=0}^{0} \left(\phi e^{-i\lambda}\right)^h + \sum_{h=1}^{\infty} \left(\phi e^{-i\lambda}\right)^h \\
&= \sum_{h=1}^{\infty} \left(\phi e^{i\lambda}\right)^h + 1 + \sum_{h=1}^{\infty} \left(\phi e^{-i\lambda}\right)^h \\
&= \frac{\phi e^{i\lambda}}{1-\phi e^{i\lambda}} + 1 + \frac{\phi e^{-i\lambda}}{1-\phi e^{-i\lambda}} \\
&= \frac{1-\phi^2}{(1-\phi e^{i\lambda})(1-\phi e^{-i\lambda})} \\
&= \frac{1-\phi^2}{1+\phi^2-2\phi\cos\lambda}.
\end{aligned}$$