STATISTICAL ANALYSIS AND MODELLING

David A. Stephens

Department of Mathematics, Imperial College

d.stephens@imperial.ac.uk

`stats.ma.ic.ac.uk/~das01/EPSCourse/`

25th March 2004

SECTION 1.
TIME SERIES ANALYSIS

1.1 INTRODUCTION

Times series analysis could be described as a branch of **applied stochastic processes**. We start with an indexed family of **random variables**

$$\{X_t : \quad t \in T\}$$

where t is the index, here taken to be time (but it could be space). T is called the index set. We have a state space of values of X.

In addition X could be **univariate** or **multivariate**. We shall concentrate on discrete time. Samples are taken at equal intervals. We wish to use time series analysis to characterize time series and understand structure.

Possibilities

State (possible values of X)	Time	Notation
Continuous	Continuous	$X(t)$
Continuous	Discrete	X_t
Discrete	Continuous	
Discrete	Discrete	

Examples: Figures 1–4, in all cases points are joined for clarity.

1. wind speed in a certain direction at a location, measured every 0.025s.
2. monthly average measurements of the flow of water in the Willamette River at Salem, Oregon.
3. the daily record of the change in average daily frequency that tells us how well an atomic clock keeps time on a day to day basis.
4. the change in the level of ambient noise in the ocean from one second to the next.
5. part of the Epstein-Barr Virus DNA sequence (the entire sequence consists of approximately 172,000 base pairs).
6. daily US Dollar/Sterling exchange rate and the corresponding returns from 1981 to 1985.

The visual appearances of these datasets are quite different. For example, consider the wind speed and atomic clock data,

- Wind speed: Adjacent points are close in value
- Atomic clock: Positive values often followed by negative values

For the numerical data, we can illustrate this using lag 1 scatter plots. Realizations of the series denoted $x_1, \ldots, x_N$. So plot x_t versus x_{t+1} as t varies from 1 to $N-1$. From these scatter plots we note the following:

- for the wind speed and US dollar series, the values are positively correlated.
- Willamette river data is similar, but points are more spread out.
- for the atomic clock data, the values are negatively correlated.
- for the ocean noise data and the US dollar returns series there is no clear clustering tendency.

We could similarly create **lag** k scatter plots by plotting x_t versus x_{t+k}, but they are unwieldy. Suppose we make the assumption that a **linear** relationship holds approximately between x_{t+k} and x_t for all k, *i.e.*,

$$x_{t+k} = \alpha_k + \beta_k x_t + \varepsilon_{t+k}$$

where ε_{t+k} is an random error term.

We can use as a summary statistic a measure of the strength of the linear relationship between two variables $\{y_t\}$ and $\{z_t\}$ say, namely the **Pearson product moment correlation coefficient**

$$\hat{\rho} = \frac{\sum(y_t - \bar{y})(z_t - \bar{z})}{\sqrt{\sum(y_t - \bar{y})^2 \sum(z_t - \bar{z})^2}}$$

where $\bar{y}$ and $\bar{z}$ are the sample means.

Hence if $y_t = x_{t+k}$ and $z_t = x_t$ we are led to the lag k sample autocorrelation for a time series:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{N-k} (x_{t+k} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}$$

with $\hat{\rho}_0 = 1$.

The sequence $\{\hat{\rho}_k\}$ is called the **sample autocorrelation sequence** (sample acs) for the time series. The sample acs for each of our time series are given in Figs. 6 and 7. Note that for the Willamette river data x_t and x_{t+6} are **negatively** correlated, while x_t and x_{t+12} are **positively** correlated (consistent with the river flow varying with a period of roughly 12 months).

The series $x_1, \ldots, x_N$ can be regarded as a realization of the corresponding random variables $X_1, \ldots, X_N$, $\hat{\rho}_k$ is an estimate of a corresponding population quantity called the lag k theoretical autocorrelation, defined as

$$\rho_k = \frac{E\left[(X_t - \mu)(X_{t+k} - \mu)\right]}{\sigma^2}$$

where $E\,[]$ is the expectation operator,

$$\mu = E\left[X_t\right]$$

is the population mean, and

$$\sigma^2 = E\left[(X_t - \mu)^2\right]$$

is the corresponding population variance.

Note that ρ_k, μ and σ^2 do not depend on t

1.2 Real-Valued discrete time processes

Denote the process by $\{X_t\}$. For fixed t, X_t is a random variable (r.v.), and hence there is an associated cumulative distribution function (cdf):

$$F_t(a) = P(X_t \le a),$$

and

$$E[X_t] = \int_{-\infty}^{\infty} x \, dF_t(x) \equiv \mu_t \qquad Var[X_t] = \int_{-\infty}^{\infty} (x - \mu_t)^2 \, dF_t(x).$$

But we are interested in the relationships between the various r.v.s that form the process. For example, for any t_1 and $t_2 \in T$,

$$F_{t_1,t_2}(a_1, a_2) = P(X_{t_1} \le a_1, X_{t_2} \le a_2)$$

gives the bivariate cdf. More generally for any $t_1, t_2, \dots, t_n \in T$,

$$F_{t_1,t_2,\dots,t_n}(a_1, a_2, \dots, a_n) = P(X_{t_1} \le a_1, \dots, X_{t_n} \le a_n)$$

1.3 STATIONARITY

We consider the subclass of **stationary processes**.

COMPLETE/STRONG/STRICT stationarity
$\{X_t\}$ is said to be completely stationary if, for all $n \geq 1$, for any

$$t_1, t_2, \ldots, t_n \in T$$

and for any τ such that

$$t_1 + \tau, t_2 + \tau, \ldots, t_n + \tau \in T$$

are also contained in the index set, the joint cdf of $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$ is the same as that of $\{X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau}\}$ i.e.,

$$F_{t_1,t_2,\ldots,t_n}(a_1, a_2, \ldots, a_n) = F_{t_1+\tau,t_2+\tau,\ldots,t_n+\tau}(a_1, a_2, \ldots, a_n),$$

so that the probabilistic structure of a completely stationary process is invariant under a shift in time.

SECOND-ORDER/WEAK/COVARIANCE stationarity

$\{X_t\}$ is said to be second-order stationary if, for all $n \geq 1$, for any

$$t_1, t_2, \ldots, t_n \in T$$

and for any τ such that $t_1 + \tau, t_2 + \tau, \ldots, t_n + \tau \in T$ are also contained in the index set, all the joint moments of orders 1 and 2 of $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$ exist and are finite. Most importantly, these moments are identical to the corresponding joint moments of $\{X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau}\}$. Hence,

$$E[X_t] \equiv \mu \qquad Var[X_t] \equiv \sigma^2 \qquad (= E\left[X_t^2\right] - \mu^2),$$

are constants independent of t. If we let $\tau = -t_1$,

$$E[X_{t_1} X_{t_2}] = E[X_{t_1+\tau} X_{t_2+\tau}] = E[X_0 X_{t_2-t_1}],$$

and with $\tau = -t_2$,

$$E[X_{t_1} X_{t_2}] = E[X_{t_1+\tau} X_{t_2+\tau}] = E[X_{t_1-t_2} X_0].$$

Hence, $E\{X_{t_1} X_{t_2}\}$ is a function of the absolute difference $|t_2 - t_1|$ only, similarly, for the **covariance** between X_{t_1} & X_{t_2}:

$$\begin{aligned} Cov\,[X_{t_1}, X_{t_2}] &= E\,[(X_{t_1} - \mu)(X_{t_2} - \mu)] \\ &= E\,[X_{t_1} X_{t_2}] - \mu^2. \end{aligned}$$

For a discrete time second-order stationary process $\{X_t\}$ we define the **autocovariance sequence** (**acvs**) by

$$\begin{aligned} s_\tau &\equiv Cov\,[X_t, X_{t+\tau}] \\ &= Cov\,[X_0, X_\tau]\,. \end{aligned}$$

NOTES:

1. τ is called the lag.
2. $s_0 = \sigma^2$ and $s_{-\tau} = s_\tau$.
3. The autocorrelation sequence (acs) is given by
$$\rho_\tau = \frac{s_\tau}{s_0} = \frac{Cov\left[X_t, X_{t+\tau}\right]}{\sigma^2}.$$
4. Since ρ_τ is a correlation coefficient, $|s_\tau| \leq s_0$.
5. The sequence $\{s_\tau\}$ is positive semidefinite, i.e., for all $n \geq 1$, for any $t_1, t_2, \ldots, t_n$ contained in the index set, and for any set of nonzero real numbers $a_1, a_2, \ldots, a_n$
$$\sum_{j=1}^{n} \sum_{k=1}^{n} s_{t_j - t_k} a_j a_k \geq 0.$$

To see this, let

$$\mathbf{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}, \quad \mathbf{V} = (X_{t_1}, X_{t_2}, \ldots, X_{t_n})^{\mathrm{T}}$$

and let Σ be the variance-covariance matrix of $\mathbf{V}$. Its j, k-th element is given by

$$s_{t_j - t_k} = E\left[(X_{t_j} - \mu)(X_{t_k} - \mu)\right]$$

Define the r.v.

$$w = \sum_{j=1}^{n} a_j X_{t_j} = \mathbf{a}^{\mathrm{T}}\mathbf{V},$$

Then

$$0 \leq Var\left[w\right] = Var\left[\mathbf{a}^{\mathrm{T}}\mathbf{V}\right] = \mathbf{a}^{\mathrm{T}} Var\left[\mathbf{V}\right]\mathbf{a} = \mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a} = \sum_{j=1}^{n}\sum_{k=1}^{n} s_{t_j - t_k} a_j a_k.$$

6. The variance-covariance matrix of equispaced X's, $(X_1, X_2, \ldots, X_N)^{\mathrm{T}}$ has the form

$$\begin{bmatrix} s_0 & s_1 & \cdots & s_{N-2} & s_{N-1} \\ s_1 & s_0 & \cdots & s_{N-3} & s_{N-2} \\ \vdots & & \ddots & & \\ s_{N-2} & s_{N-3} & \cdots & s_0 & s_1 \\ s_{N-1} & s_{N-2} & \cdots & s_1 & s_0 \end{bmatrix}$$

which is known as a symmetric Toeplitz matrix – all elements on a diagonal are the same.

7. Note the above matrix has only N unique elements, $s_0, s_1, \ldots, s_{N-1}$.

8. A stochastic process $\{X_t\}$ is called Gaussian if, for all $n \geq 1$ and for any $t_1, t_2, \ldots, t_n$ contained in the index set, the joint cdf of $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ is multivariate Gaussian.

   - 2nd-order stationary Gaussian $\Rightarrow$ complete stationarity
     - follows as the multivariate Normal distribution is completely characterized by 1st and 2nd moments
     - not true in general.
   - Complete stationarity $\Rightarrow$ 2nd-order stationary in general.

1.4 DISCRETE STATIONARY PROCESSES

Example 1.4.1 White noise process
Also known as a purely random process. Let $\{X_t\}$ be a sequence of uncorrelated r.v.s such that

$$E[X_t] = \mu \qquad Var[X_t] = \sigma^2 \qquad \forall t$$

and

$$s_\tau = \begin{cases} \sigma^2 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases} \qquad \text{or} \qquad \rho_\tau = \begin{cases} 1 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases}$$

forms a basic building block in time series analysis. Very different realizations of white noise can be obtained for different distributions of $\{X_t\}$. Examples are given in Figures 8 and 9 for processes with (a) Gaussian, (b) exponential, (c) uniform and (d) truncated Cauchy distributions.

Example 1.4.2 q-th order moving average process MA(q)
X_t can be expressed in the form

$$X_t = \mu - \theta_{0,q}\epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q} = \mu - \sum_{j=0}^{q} \theta_{j,q}\epsilon_{t-j},$$

where μ and $\theta_{j,q}$'s are constants ($\theta_{0,q} \equiv -1, \theta_{q,q} \neq 0$), and $\{\epsilon_t\}$ is a zero-mean white noise process with variance σ_ϵ^2.
We assume $E[X_t] = \mu = 0$. Then

$$Cov[X_t, X_{t+\tau}] = E\{X_t X_{t+\tau}\}$$

Recall: $Cov(X,Y) = E\{(X - E\{X\})(Y - E\{Y\})\}$. Since

$E\{\epsilon_t \epsilon_{t+\tau}\} = 0 \;\; \forall\, \tau \neq 0$ we have for $\tau \geq 0$.

$$
\begin{aligned}
Cov\,[X_t, X_{t+\tau}] &= \sum_{j=0}^{q} \sum_{k=0}^{q} \theta_{j,q} \theta_{k,q} E\{\epsilon_{t-j} \epsilon_{t+\tau-k}\} \\
&= \sigma_\epsilon^2 \sum_{j=0}^{q-\tau} \theta_{j,q} \theta_{j+\tau,q} \qquad (k = j + \tau) \\
&\equiv s_\tau ,
\end{aligned}
$$

which does not depend on t. Since $s_\tau = s_{-\tau}$, $\{X_t\}$ is a stationary process with acvs given by

$$
s_\tau = \begin{cases} \sigma_\epsilon^2 \sum_{j=0}^{q-|\tau|} \theta_{j,q} \theta_{j+|\tau|,q} & |\tau| \leq q \\ 0 & |\tau| > q \end{cases}
$$

N.B. No restrictions were placed on the $\theta_{j,q}$'s to ensure stationarity. (Though obviously, $|\theta_{j,q}| < \infty \;\; \forall\, j$).

Examples (Figures 10 and 11)

$$X_t = \epsilon_t - \theta_{1,1}\epsilon_{t-1} \qquad \text{MA(1)}$$

$\boxed{\text{acvs:}}$

$$s_\tau = \sigma_\epsilon^2 \sum_{j=0}^{1-|\tau|} \theta_{j,1}\theta_{j+|\tau|,1} \quad |\tau| \leq 1,$$

so,

$$s_0 = \sigma_\epsilon^2(\theta_{0,1}\theta_{0,1} + \theta_{1,1}\theta_{1,1}) = \sigma_\epsilon^2(1 + \theta_{1,1}^2);$$

and,

$$s_1 = \sigma_\epsilon^2\theta_{0,1}\theta_{1,1} = -\sigma_\epsilon^2\theta_{1,1}.$$

$\boxed{\text{acs:}}$

$$\rho_\tau = \frac{s_\tau}{s_0} : \rho_0 = 1.0 \qquad \rho_1 = \frac{-\theta_{1,1}}{1+\theta_{1,1}^2}$$

For $\theta_{1,1} = 1.0, \sigma_\epsilon^2 = 1.0$, we have,

$$s_0 = 2.0, s_1 = -1.0, s_2, s_3, \ldots = 0.0,$$

giving,

$$\rho_0 = 1.0, \rho_1 = -0.5, \rho_2, \rho_3, \ldots = 0.0.$$

For $\theta_{1,1} = -1.0, \sigma_\epsilon^2 = 1.0$, we have,

$$s_0 = 2.0, s_1 = 1.0, s_2, s_3, \ldots = 0.0,$$

giving,

$$\rho_0 = 1.0, \rho_1 = 0.5, \rho_2, \rho_3, \ldots = 0.0.$$

Note: if we replace $\theta_{1,1}$ by $\theta_{1,1}^{-1}$ the model becomes

$$X_t = \epsilon_t - \frac{1}{\theta_{1,1}} \epsilon_{t-1}$$

and the autocorrelation becomes

$$\rho_1 = \frac{-\dfrac{1}{\theta_{1,1}}}{1 + \left(\dfrac{1}{\theta_{1,1}}\right)^2} = \frac{-\theta_{1,1}}{\theta_{1,1}^2 + 1},$$

i.e., is unchanged. Thus we cannot identify the MA(1) process uniquely from the autocorrelation.

Example 1.4.3 p-th order autoregressive process AR(p)
$\{X_t\}$ is expressed in the form

$$X_t = \phi_{1,p} X_{t-1} + \phi_{2,p} X_{t-2} + \ldots + \phi_{p,p} X_{t-p} + \epsilon_t,$$

where $\phi_{1,p}, \phi_{2,p}, \ldots, \phi_{p,p}$ are constants ($\phi_{p,p} \neq 0$) and $\{\epsilon_t\}$ is a zero mean white noise process with variance σ_ϵ^2..

In contrast to the parameters of an MA(q) process, the $\{\phi_{k,p}\}$ **must satisfy certain conditions** for $\{X_t\}$ to be a stationary process – not all AR(p) processes are stationary

Examples (Figures 12 and 13)

$$\begin{aligned} X_t &= \phi_{1,1} X_{t-1} + \epsilon_t = \phi_{1,1}\{\phi_{1,1} X_{t-2} + \epsilon_{t-1}\} + \epsilon_t = \phi_{1,1}^2 X_{t-2} + \phi_{1,1}\epsilon_{t-1} + \epsilon_t \quad (1) \\ &\vdots \\ &= \sum_{k=0}^{\infty} \phi_{1,1}^k \epsilon_{t-k} \qquad \text{(initial condition } X_{-N} = 0; \text{let } N \to \infty\text{)} \end{aligned}$$

Note $E\left[X_t\right] = 0$.

$$Var\left[X_t\right] = Var\left[\sum_{k=0}^{\infty} \phi_{1,1}^{k} \epsilon_{t-k}\right] = \sum_{k=0}^{\infty} Var\{\phi_{1,1}^{k} \epsilon_{t-k}\} = \sigma_{\epsilon}^{2} \sum_{k=0}^{\infty} \phi_{1,1}^{2k}.$$

For $Var\left[X_t\right] < \infty$ we must have $|\phi_{1,1}| < 1$, in which case

$$Var\left[X_t\right] = \frac{\sigma_{\epsilon}^{2}}{1 - \phi_{1,1}^{2}}.$$

To find the form of the acvs, we notice that for $\tau > 0$, $X_{t-\tau}$ is a linear function of $\epsilon_{t-\tau}, \epsilon_{t-\tau-1}, \ldots$ and is therefore uncorrelated with ϵ_t. Hence

$$E\left[\epsilon_t X_{t-\tau}\right] = 0.$$

Assuming stationarity and multiplying the defining equation (1) by $X_{t-\tau}$:

$$\begin{aligned} X_t X_{t-\tau} &= \phi_{1,1} X_t X_{t-\tau} + \epsilon_t X_{t-\tau} \\ \Longrightarrow E\left[X_t X_{t-\tau}\right] &= \phi_{1,1} E\left[X_{t-1} X_{t-\tau}\right] \end{aligned}$$

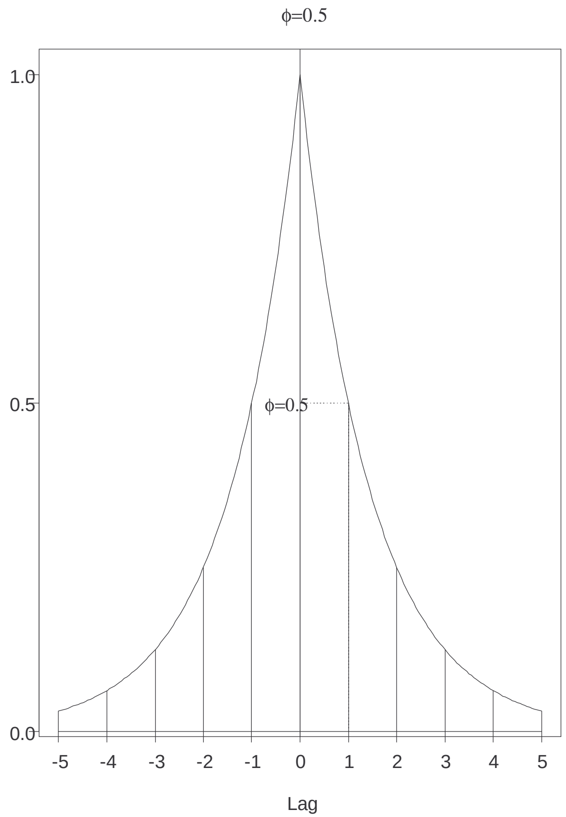
so that

$$s_\tau = \phi_{1,1} s_{\tau-1} = \phi_{1,1}^2 s_{\tau-2} = \ldots = \phi_{1,1}^\tau s_0 \qquad \Rightarrow \rho_\tau = \frac{s_\tau}{s_0} = \phi_{1,1}^\tau$$

However ρ_τ is an even function of τ, so

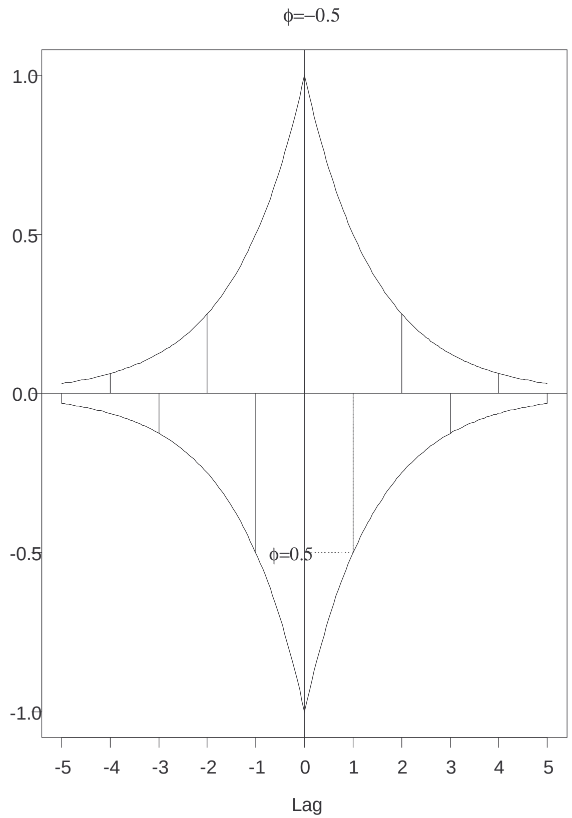
$$\rho_\tau = \phi_{1,1}^{|\tau|} \qquad \tau = 0, \pm 1, \pm 2, \ldots .$$

giving exponential decay



$$\phi_{1,1} = 0.5$$

•



$$\phi_{1,1} = -0.5$$

Example 1.4.4 (p,q)**'th order autoregressive-moving average process ARMA**(p,q)

Here $\{X_t\}$ is expressed as

$$X_t = \phi_{1,p}X_{t-1} + \ldots + \phi_{p,p}X_{t-p} + \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q},$$

where the $\phi_{j,p}$'s and the $\theta_{j,q}$'s are all constants ($\phi_{p,p} \neq 0; \theta_{q,q} \neq 0$) and again $\{\epsilon_t\}$ is a zero mean white noise process with variance σ_ϵ^2.

The ARMA class is important as many data sets may be approximated in a more parsimonious way (meaning fewer parameters are needed) by a mixed ARMA model than by a pure AR or MA process.

1.5 MODELS FOR CHANGING VARIANCE

Objective: obtain better estimates of local variance in order to obtain a better assessment of risk.(for example, in finance)

Example 1.5.1 p'th order ARCH(p)

ARCH stands for **autoregressive conditionally heteroscedastic**

Assume we have a derived time series $\{Y_t\}$ that is (approximately) uncorrelated but has a variance (volatility) that changes through time,

$$Y_t = \sigma_t \varepsilon_t \tag{2}$$

where $\{\varepsilon_t\}$ is a white noise sequence with zero mean and unit variance.

Here, σ_t represents the local conditional standard deviation of the process. Note that σ_t is not observable directly.

$\{Y_t\}$ is ARCH(p) if it satisfies equation (2) and

$$\sigma_t^2 = \alpha + \beta_{1,p} y_{t-1}^2 + \ldots + \beta_{p,p} y_{t-p}^2, \qquad (3)$$

where $\alpha > 0$ and $\beta_{j,p} \geq 0, j = 1, \ldots, p$ (to ensure the variance remains positive), and y_{t-1} is the observed value of the derived time series at time $(t-1)$

Notes:

(a) the absence of the error term in equation (3).

(b) unconstrained estimation often leads to violation of the non-negativity constraints that are needed to ensure positive variance.

(c) quadratic form (i.e. modelling σ_t^2) prevents modelling of asymmetry in volatility (i.e. volatility tends to be higher after a decrease than after an equal increase and ARCH cannot account for this).

Example 1.5.2 ARCH(1)

$$\sigma_t^2 = \alpha + \beta_{1,1} y_{t-1}^2$$

Define, $v_t = y_t^2 - \sigma_t^2 \Rightarrow \sigma_t^2 = y_t^2 - v_t$. The model can also be written:

$$y_t^2 = \alpha + \beta_{1,1} y_{t-1}^2 + v_t,$$

i.e. an AR(1) model for $\{y_t^2\}$ where the errors, $\{v_t\}$, have zero mean, but as $v_t = \sigma_t^2(\epsilon_t^2 - 1)$ the errors are heteroscedastic.

Example 1.5.3 (p, q)**'th order generalized autoregressive conditionally heteroscedastic model GARCH**(p, q)
$\{Y_t\}$ is GARCH(p, q) if it satisfies equation (2) and

$$\sigma_t^2 = \alpha + \beta_{1,p} y_{t-1}^2 + \ldots + \beta_{p,p} y_{t-p}^2 + \gamma_{1,q} \sigma_{t-1}^2 + \ldots \gamma_{q,q} \sigma_{t-q}^2,$$

where the parameters are chosen to ensure positive variance.

Example 1.5.4 Stochastic volatility models SV
Stochastic volatility models treat σ_t as an unobserved random variable which is assumed to follow a certain stochastic process. The specification for the derived series $\{Y_t\}$ is:

$$Y_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \exp(h_t),$$

where ε_t is white noise with zero mean and unit variance, and let h_t, for example, be an AR(1) process:

$$h_t = \alpha + \beta_{1,1} h_{t-1} + \eta_t,$$

where $\{\eta_t\}$ is a white noise process with variance σ_η^2.

If $|\beta_{1,1}| < 1$, h_t is stationary $\Rightarrow$ Y_t stationary.

Notes:
(a) unlike the GARCH specification, h_t (which defines in turn σ_t) is NOT deterministic.

(b) the exponential specification ensures positive conditional variance.

(c) can be further generalized by assuming, for example, h_t follows an ARMA(p, q) model.

Example 1.5.5 Harmonic with additive white noise (see Figure 14)
Here $\{X_t\}$ is expressed as

$$X_t = \cos(2\pi f_0 t + \phi) + \epsilon_t$$

f_0 is a fixed frequency and $\{\epsilon_t\}$ is zero mean white noise with variance σ_ϵ^2.

Case (a) ϕ is constant.

$$E\left[X_t\right] = E\left[\cos(2\pi f_0 t + \phi)\right] + E\left[\epsilon_t\right] = \cos(2\pi f_0 t + \phi).$$

so, mean depends on $t \Rightarrow$ not stationary.

Case (b): $\phi \sim U[-\pi, \pi]$ and independent. of $\{\epsilon_t\}$.

$$E\left[X_t\right] = E\left[\cos(2\pi f_0 t + \phi) + \epsilon_t\right] = E\{\cos(2\pi f_0 t + \phi)\}$$

Now,

$$E\{\cos(2\pi f_0 t + \phi)\} = \int_{-\pi}^{\pi} \cos(2\pi f_0 t + \phi) \frac{1}{2\pi}\, d\phi = \left[\frac{\sin(2\pi f_0 t + \phi)}{2\pi}\right]_{-\pi}^{\pi} = 0.$$

So $E\left[X_t\right] = 0$, and, using the fact that $\{e_t\}$ and ϕ are independent.

$$\begin{aligned} Cov\left[X_t, X_{t+\tau}\right] &= E\left[X_t X_{t+\tau}\right] \\ &= E\left[\left[\cos(2\pi f_0 t + \phi) + \epsilon_t\right]\left[\cos(2\pi f_0 (t+\tau) + \phi) + \epsilon_{t+\tau}\right]\right] \\ &= E\left[\cos(2\pi f_0 t + \phi)\cos(2\pi f_0 t + \phi + 2\pi f_0 \tau)\right] + E\left[\epsilon_t \epsilon_{t+\tau}\right]. \end{aligned}$$

Recall, as $\{\epsilon_t\}$ is white noise we have,

$$E\{\epsilon_t \epsilon_{t+\tau}\} = \begin{cases} \sigma_\epsilon^2 & \text{if } \tau = 0, \\ 0 & \text{if } \tau \neq 0, \end{cases}$$

So, for $\tau = 0$,

$$Cov\{X_t, X_t\} = s_0 = E\{\cos^2(2\pi f_0 t + \phi)\} + \sigma_\epsilon^2.$$

Now,

$$\begin{aligned} E\{\cos^2(2\pi f_0 t + \phi)\} &= \int_{-\pi}^{\pi} \cos^2(2\pi f_0 t + \phi) \frac{1}{2\pi}\, d\phi \\ &= \frac{1}{2} \int_{-\pi}^{\pi} [1 + \cos(4\pi f_0 t + 2\phi)] \frac{1}{2\pi}\, d\phi = \frac{1}{2}. \end{aligned}$$

So, $s_0 = \frac{1}{2} + \sigma_\epsilon^2$, and for $\tau > 0$,

$$\begin{aligned} Cov\left[X_t, X_{t+\tau}\right] &= s_\tau = E\left[\cos(2\pi f_0 t + \phi)\cos(2\pi f_0 t + \phi + 2\pi f_0 \tau)\right] \\ &= \frac{1}{2} E\left[\cos(4\pi f_0 t + 2\phi + 2\pi f_0 \tau) + \cos(2\pi f_0 \tau)\right] \\ &= \frac{1}{2}\int_{-\pi}^{\pi} \cos(2\pi f_0 \tau) \frac{1}{2\pi}\, \mathrm{d}\phi \\ &= \frac{\cos(2\pi f_0 \tau)}{2}\left[\frac{\phi}{2\pi}\right]_{-\pi}^{\pi} = \frac{\cos(2\pi f_0 \tau)}{2} \end{aligned}$$

which does not depend on $t \Rightarrow X_t$ is stationary.

1.5.1 Trend removal and seasonal adjustment

There are certain, quite common, situations where the observations exhibit a trend – a tendency to increase or decrease slowly steadily over time – or may fluctuate in a periodic manner due to seasonal effects. The model is modified to

$$X_t = \mu_t + Y_t$$

- μ_t = time dependent mean.
- Y_t = zero mean stationary process.

Example 1.5.6 Trend adjustment for CO^2 data: $\{X_t\}$ is monthly atmospheric CO^2 concentrations expressed in parts per million (ppm) derived from in situ air samples collected at Mauna Loa observatory, Hawaii. Monthly data from May 1988 – December 1998, giving $N = 128$. The data are plotted in Figure 14. Model suggested by plot:

$$X_t = \alpha + \beta t + Y_t.$$

(a) Estimate α and β by least squares, and work with the residuals

$$\hat{Y}_t = X_t - \hat{\alpha} - \hat{\beta} t.$$

For the CO^2 data these are shown in the middle plot of figure 14.

(b) Take first differences: for the CO^2 data these are shown in the bottom plot of figure 14.

$$X_t^{(1)} = X_t - X_{t-1} = \alpha + \beta t + Y_t - (\alpha + \beta(t-1) + Y_{t-1}) = \beta + Y_t - Y_{t-1}.$$

Note: if $\{Y_t\}$ is stationary so is $\{Y_t^{(1)}\}$ In the case of linear trend, if we difference again:

$$\begin{aligned} X_t^{(2)} &= X_t^{(1)} - X_{t-1}^{(1)} = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= (\beta + Y_t - Y_{t-1}) - (\beta + Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2}, \quad (\equiv Y_t^{(1)} - Y_{t-1}^{(1)} = Y_t^{(2)}), \end{aligned}$$

so that the effect of $\mu_t(= \alpha + \beta t)$ has been completely removed.

If μ_t is a polynomial of degree $(d-1)$ in t, then dth differences of μ_t will be zero ($d = 2$ for linear trend). Further,

$$X_t^{(d)} = \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k} = \sum_{k=0}^{d} \binom{d}{k} (-1)^k Y_{t-k}.$$

There are other ways of writing this. Define the difference operator

$$\Delta = (1 - B)$$

where $BX_t = X_{t-1}$ is the *backward shift operator* (sometimes known as the *lag operator* L – especially in econometrics). Then,

$$X_t^{(d)} = \Delta^d X_t = \Delta^d Y_t.$$

For example, for $d = 2$:

$$
\begin{aligned}
X_t^{(2)} &= (1-B)^2 X_t = (1-B)(X_t - X_{t-1}) \\
&= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\
&= (\beta + Y_t - Y_{t-1}) - (\beta + Y_{t-1} - Y_{t-2}) \\
&= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\
&= (1-B)^2 Y_t = \Delta^2 Y_t.
\end{aligned}
$$

This notation can be incorporated into the ARMA set up, recall if $\{X_t\}$ is ARMA(p, q),

$$X_t = \phi_{1,p} X_{t-1} + \ldots + \phi_{p,p} X_{t-p} + \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q},$$

$$\begin{aligned} X_t - \phi_{1,p} X_{t-1} - \ldots - \phi_{p,p} X_{t-p} &= \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q} \\ (1 - \phi_{1,p} B - \phi_{2,p} B^2 - \ldots - \phi_{p,p} B^p) X_t &= (1 - \theta_{1,q} B - \theta_{2,q} B^2 - \ldots - \theta_{q,q} B^q)\epsilon_t \\ &. \end{aligned}$$

We write this

$$\Phi(B)X_t = \Theta(B)\epsilon_t$$

where

$$\begin{aligned}\Phi(B) &= 1 - \phi_{1,p}B - \phi_{2,p}B^2 - \ldots - \phi_{p,p}B^p \\ \Theta(B) &= 1 - \theta_{1,q}B - \theta_{2,q}B^2 - \ldots - \theta_{q,q}B^q\end{aligned}$$

are known as the *associated* or *characteristic polynomials*.

Further, we can generalize the class of ARMA models to include differencing to account for certain types of non-stationarity, namely,

- X_t is called **ARIMA**(p, d, q) if

$$\begin{aligned} \Phi(B)(1-B)^d X_t &= \Theta(B)\epsilon_t, \\ \Phi(B)\Delta^d X_t &= \Theta(B)\epsilon_t. \end{aligned}$$

Seasonal adjustment

The model is modified to

$$X_t = s_t + Y_t$$

where

- $\{s_t\}$ is the **seasonal** component,
- $\{Y_t\}$ is zero mean **stationary** process.

Presuming that the seasonal component maintains a constant pattern over time with period s, there are again several approaches to removing s_t. A popular approach used by Box & Jenkins is to use the operator $(1 - B^s)$:

$$\begin{aligned} X_t^{(s)} &= (1 - B^s)X_t = X_t - X_{t-s} \\ &= (s_t + Y_t) - (s_{t-s} + Y_{t-s}) \\ &= Y_t - Y_{t-s} \end{aligned}$$

since s_t has period s (and so $s_{t-s} = s_t$).

Figure 16 shows this technique applied to the CO^2 data – most of the seasonal structure and trend has been removed by applying the following differencing:

$$(1 - B^s)(1 - B)X_t.$$

1.6 THE GENERAL LINEAR PROCESS

Consider a process of the form

$$X_t = \sum_{k=-\infty}^{\infty} g_k \epsilon_{t-k},$$

where $\{\epsilon_t\}$ is a purely random process, and $\{g_k\}$ is a given sequence of constants satisfying $\sum_{k=\infty}^{-\infty} g_k^2 < \infty$.

This condition ensures that $\{X_t\}$ has finite variance. Now we know $|\rho_t| \leq 1$, so

$$|s_\tau| = |Cov\,[X_t, X_{t-\tau}]| \leq \sigma_X^2 = \sigma_\epsilon^2 \sum_k g_k^2 < \infty.$$

so the covariance is bounded also.

If

$$g_{-1}, g_{-2}, \ldots = 0,$$

then we obtain what is called the General Linear Process

$$X_t = \sum_{k=0}^{\infty} g_k \epsilon_{t-k},$$

where X_t depends only on past and present values $\epsilon_t, \epsilon_{t-2}, \epsilon_{t-2}, \ldots$ of the purely random process. Consider the function

$$G(z) = \sum_{k=0}^{\infty} g_k z^k,$$

"z-polynomial" where $z = e^{-i\omega}$. Note $X_t = G(B)\epsilon_t$.

Then write

$$G(z) = \frac{G_1(z)}{G_2(z)}$$

Call the zeros of $G_2(z)$ (the "poles" of $G(z)$) in the **complex plane** $z_1, z_2, \ldots, z_p$, where the zeros are ordered so that $z_1, \ldots, z_k$ are inside and $z_{k+1}, \ldots, z_p$ are outside the unit circle $|z| = 1$. Then,

$$\begin{aligned}
\frac{1}{G_2(z)} = \sum_{j=1}^{p} \frac{A_j}{z - z_j} \quad &= \quad \left(\sum_{j=1}^{k} \frac{A_j}{z} \times \frac{1}{\left(1 - \frac{z_j}{z}\right)} \right) + \left(\sum_{j=k+1}^{p} \frac{A_j}{z_j} \times \frac{-1}{\left(1 - \frac{z}{z_j}\right)} \right) \\
&= \quad \sum_{j=1}^{k} \frac{A_j}{z} \sum_{l=0}^{\infty} \left(\frac{z_j}{z}\right)^l - \sum_{j=k+1}^{p} \frac{A_j}{z_j} \sum_{l=0}^{\infty} \left(\frac{z}{z_j}\right)^l
\end{aligned}$$

which is convergent for $|z| = 1$.

Replace z by the backshift operator B and apply to $\{\epsilon_t\}$:

$$\begin{aligned}
\left\{\frac{1}{G_2(B)}\right\}\epsilon_t &= \left\{\sum_{j=1}^{k} A_j B^{-1}\sum_{l=0}^{\infty} z_j^l B^{-l} - \sum_{j=k+1}^{p} A_j z_j^{-1}\sum_{l=0}^{\infty} z_j^{-l} B^l\right\}\epsilon_t \\
&= \sum_{j=1}^{k} A_j \sum_{l=0}^{\infty} z_j^l \epsilon_{t+l+1} - \sum_{j=k+1}^{p} A_j \sum_{l=0}^{\infty} \underbrace{z_j^{-l-1}}_{\text{outside}} \epsilon_{t-l}.
\end{aligned}$$

Hence, if all the roots of $G_2(z)$ are outside the unit circle (i.e. all the poles of $G(z)$ are outside the unit circle) only past and present values of $\{\epsilon_t\}$ are involved and the General Linear Process exists.

Another way of stating this is that

$$G(z) < \infty \qquad |z| \leq 1$$

i.e., $G(z)$ is analytic inside and on the unit circle. Thus

- all the poles of $G(z)$ lie outside the unit circle
- all the roots of $G^{-1}(z) = 0$ lie outside the unit circle

Consider the MA(q) model

$$X_t = \Theta(B)\epsilon_t,$$

then,

$$\Theta^{-1}(B)X_t = \epsilon_t$$

and in general, the expansion of $\Theta^{-1}(B)$ is a polynomial of infinite order. Similarly, consider the AR(p) model

$$\Phi(B)X_t = \epsilon_t,$$

then,

$$X_t = \Phi^{-1}(B)\epsilon_t$$

Hence

$$\begin{array}{lcl} \text{MA (finite order)} & \equiv & \text{AR (infinite order)} \\ \text{AR (finite order)} & \equiv & \text{MA (infinite order)} \end{array}$$

provided the infinite order expansions exist

1.6.1 Invertibility

Consider inverting the general linear process into autoregressive form

$$\begin{aligned} X_t &= \sum_{k=0}^{\infty} g_k \epsilon_{t-k} = \sum_{k=0}^{\infty} g_k B^k \epsilon_t \\ &= G(B)\epsilon_t \end{aligned}$$

so that

$$G^{-1}(B)X_t = \epsilon_t$$

The expansion of $G^{-1}(B)$ in powers of B gives the required autoregressive form **provided** $G^{-1}(B)$ **admits a power series expansion**

$$G^{-1}(z) = \sum_{k=0}^{\infty} h_k z^k$$

i.e. if $G^{-1}(z)$ is analytic, $|z| \leq 1$.

Thus the model is invertible if All the poles of $G^{-1}(z)$ are outside the unit circle.

$$G^{-1}(z) < \infty, \quad |z| \leq 1.$$

For the MA(q) process, $G(z) = \Theta(z)$, and so the invertibility condition is that $\Theta(z)$ has no roots inside or on the unit circle; i.e. all the roots of $\Theta(z)$ lie outside the unit circle.

Example 1.6.1 Consider the following process

$$X_t = \epsilon_t - 1.3\epsilon_{t-1} + 0.4\epsilon_{t-2} \Longrightarrow X_t = (1 - 1.3B + 0.4B^2)\epsilon_t = \Theta(B)\epsilon_t$$

to check if invertible, find roots of $\Theta(z) = 1 - 1.3z + 0.4z^2$,

$$1 - 1.3z + 0.4z^2 = 0 \Longrightarrow 4z^2 - 13z + 10 = 0 \Longrightarrow (4z - 5)(z - 2) = 0$$

roots of $\Theta(z)$ are $z = 2$ and $z = 5/4$, which are both outside the unit circle $\Rightarrow$ invertible.

1.6.2 Stationarity

For the AR(p) process

$$\Phi(B)X_t = \epsilon_t$$

so that

$$X_t = \Phi^{-1}(B)\epsilon_t = G(B)\epsilon_t,$$

so that $G(z) = \Phi^{-1}(z)$. Hence the requirement for stationarity is that all the roots of $G^{-1}(z) = \Phi(z)$ must lie outside the unit circle.
For the MA(q) process

$$X_t = \Theta(B)\epsilon_t = G(B)\epsilon_t$$

and since $G(B) = \Theta(B)$ is a polynomial of finite order $G(z) < \infty$, $|z| \leq 1$, automatically.

Example 1.6.2 Determine whether the following model is stationary and/or invertible,

$$X_t = 1.3X_{t-1} - 0.4X_{t-2} + \epsilon_t - 1.5\epsilon_{t-1}.$$

Writing in B notation:

$$(1 - 1.3B + 0.4B^2)X_t = (1 - 1.5B)\epsilon_t$$

we have

$$\Phi(z) = 1 - 1.3z + 0.4z^2$$

with roots $z = 2$ and $5/4$ (from previous example), so the roots of $\Phi(z) = 0$ both lie outside the unit circle, therefore model is stationary, and

$$\Theta(z) = 1 - 1.5z,$$

so the root of $\Theta(z) = 0$ is given by $z = 2/3$ which lies inside the unit circle and the model is not invertible.

1.6.3 Directionality and Reversibility

Consider again the general linear model

$$X_t = \sum_{k=0}^{\infty} g_k \epsilon_{t-k} = \sum_{k=0}^{\infty} g_k B^k \epsilon_t = G(B)\epsilon_t$$

The reversed form is clearly,

$$X_t = \sum_{k=0}^{\infty} g_k \epsilon_{t+k} = \sum_{k=0}^{\infty} g_k B^{-k} \epsilon_t = G\left(\frac{1}{B}\right)\epsilon_t,$$

with some stationarity condition.

Now consider the ARMA(p, q) model given by

$$\Phi(B)X_t = \Theta(B)\epsilon_t,$$

where,

$$\begin{aligned}
\Phi(B) &= 1 - \phi_{1,p}B - \phi_{2,p}B^2 - \ldots - \phi_{p,p}B^p \\
\Theta(B) &= 1 - \theta_{1,q}B - \theta_{2,q}B^2 - \ldots - \theta_{q,q}B^q
\end{aligned}$$

The reversed form of the ARMA(p, q) model is,

$$\Phi\left(\frac{1}{B}\right)X_t = \Theta\left(\frac{1}{B}\right)\epsilon_t \Longrightarrow \Phi^R(B)X_t = B^{p-q}\Theta^R\epsilon_t$$

where,

$$\begin{aligned}
\Phi^R(B) &= B^p - \phi_{1,p}B^{p-1} - \phi_{2,p}B^{p-2} - \ldots - \phi_{p,p} \\
\Theta^R(B) &= B^q - \theta_{1,q}B^{q-1} - \theta_{2,q}B^{q-2} - \ldots - \theta_{q,q}
\end{aligned}$$

For example, for the ARMA(1,1) model,

$$(1 - \phi_{1,1})X_t = (1 - \theta_{1,1})\epsilon_t,$$

reversed form is

$$(B - \phi_{1,1})X_t = (B - \theta_{1,1})\epsilon_t$$

Now $\Phi(z) = 1 - \phi_{1,1}z$, and a root is the solution of $1 - \phi_{1,1}z = 0$, i.e.,

$$|z| = \left|\frac{1}{\phi_{1,1}}\right| > 1 \Rightarrow |\phi_{1,1}| < 1.$$

But, $\Phi^R(z) = z - \phi_{1,1}$, and so a root is the solution of $z - \phi_{1,1} = 0$, i.e., $z = \phi_{1,1}$. But, since for stationarity $|\phi_{1,1}| < 1$ we have

$$|z| = |\phi_{1,1}| < 1,$$

so the root of $\Phi^R(z)$ is inside the unit circle.

Hence the standard assumption for stationarity (roots outside the unit circle) has within it an assumption of directionality. [N.B. only if the roots of $\Phi(z)$ are on the unit circle is model ALWAYS non-stationary].

SECTION 2.

Spectral Representations

Spectral analysis is a study of the frequency domain characteristics of a process, and describes the contribution of each frequency to the variance of the process. Let us define a complex "jump" process $\{Z(f)\}$ on the interval $[0, 1/2]$, such that

$$dZ(f) \equiv \begin{cases} Z(f + df) - Z(f), & 0 \leq f < 1/2; \\ 0, & f = 1/2; \\ dZ^*(-f), & -1/2 \leq f < 0, \end{cases}$$

where df is a small positive increment. If the intervals $[f, f + df]$ and $[f', f' + df']$ are non-intersecting subintervals of $[-1/2, 1/2]$, then the r.v.'s $dZ(f)$ and $dZ(f')$ are uncorrelated.

We say that the process has **orthogonal increments**, and the process itself is called an **orthogonal process** – this orthogonality results is very important.

Let $\{X_t\}$ be a real-valued discrete time stationary process, with zero mean, the **spectral representation theorem** states that there exists such an orthogonal process $\{Z(f)\}$, defined on $(-1/2, 1/2]$, such that

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi ft}\, dZ(f)$$

for all integers t. The process $\{Z(f)\}$ has the following properties:

- $E\{dZ(f)\} = 0 \;\; \forall \;\; |f| \leq 1/2.$
- $E\{|dZ(f)|^2\} \equiv dS^{(I)}(f)$ say $\forall \;\; |f| \leq 1/2$, where $dS^{(I)}(f)$ is called the integrated spectrum of $\{X_t\}$, and

- for any two distinct frequencies f and $f' \in (-1/2, 1/2]$

$$Cov\{dZ(f'), dZ(f)\} = E\{dZ^*(f')dZ(f)\} = 0.$$

The spectral representation

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi ft}\, dZ(f) = \int_{-1/2}^{1/2} e^{i2\pi ft}\, |dZ(f)| e^{i\,\arg\{dZ(f)\}},$$

means that we can represent any discrete stationary process as an "infinite" sum of complex exponentials at frequencies f with associated random amplitudes $|dZ(f)|$ and random phases $\arg\{dZ(f)\}$.

The orthogonal increments property can be used to define the relationship between the autocovariance sequence $\{s_\tau\}$ and the integrated spectrum $S^I(f)$:

$$
\begin{aligned}
s_\tau = E\left[X_t X_{t+\tau}\right] &= E\left[X_t^* X_{t+\tau}\right] \\
&= E\left[\int_{-1/2}^{1/2} e^{-i2\pi f' t}\, dZ^*(f') \int_{-1/2}^{1/2} e^{i2\pi f(t+\tau)}\, dZ(f)\right] \\
&= \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{i2\pi(f-f')t} e^{i2\pi f\tau} E\{dZ^*(f')dZ(f)\}.
\end{aligned}
$$

Because of the orthogonal increments property,

$$E\{dZ^*(f')dZ(f)\} = \begin{cases} dS^{(I)}(f) & f = f' \\ 0 & f \neq f' \end{cases}$$

so

$$s_\tau = \int_{-1/2}^{1/2} e^{i2\pi f\tau}\, dS^{(I)}(f),$$

which shows that the integrated spectrum determines the acvs for a stationary process. If in fact $S^{(I)}(f)$ is differentiable everywhere with a derivative denoted by $S(f)$ we have

$$E\{|dZ(f)|^2\} = dS^{(I)}(f) = S(f)\, df.$$

The function $S(\cdot)$ is called the spectral density function (sdf). Hence

$$s_\tau = \int_{-1/2}^{1/2} e^{i2\pi ft} S(f)\, df.$$

But a square summable deterministic sequence $\{g_t\}$ say has the **Fourier representation**

$$g_t = \int_{-1/2}^{1/2} G(f) e^{i2\pi ft}\, df \qquad \text{where} \qquad G(f) = \sum_{t=-\infty}^{\infty} g_t e^{-i2\pi ft},$$

If we assume that $S(f)$ is square integrable, then $S(f)$ is the **Fourier transform** of $\{s_\tau\}$,

$$S(f) = \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f\tau}.$$

Hence,

$$\{s_\tau\} \longleftrightarrow S(f),$$

i.e., $\{s_\tau\}$ and $S(f)$ are a FT. pair.

2.1 SPECTRAL DENSITY FUNCTION

Subject to its existence, $S(\cdot)$ has the following interpretation: $S(f)\,df$ is the average contribution (over all realizations) to the power from components with frequencies in a small interval about f. The power – or variance – is

$$\int_{-1/2}^{1/2} S(f)\,df.$$

Hence, $S(f)$ is often called the power spectral density function or just power spectrum.

2.1.1 Properties

1. $S^{(I)}(f) = \int_{-1/2}^{f} S(f')\, df'$.

2. $0 \leq S^{(I)}(f) \leq \sigma^2$ where $\sigma^2 = Var\,[X_t]\,; \quad S(f) \geq 0$.

3. $S^{(I)}(-1/2) = 0; \quad S^{(I)}(1/2) = \sigma^2; \quad \int_{-1/2}^{1/2} S(f)\, df = \sigma^2$.

4. $f < f' \Rightarrow S^{(I)}(f) \leq S^{(I)}(f'); \quad S(-f) = S(f)$.

Except, basically, for the scaling factor $\sigma^2, S^{(I)}(f)$ has all the properties of a probability distribution function, and hence is sometimes called a spectral distribution function.

2.1.2 Classification of Spectra

For most practical purposes any integrated spectrum, $S^{(I)}(f)$ can be written as

$$S^{(I)}(f) = S_1^{(I)}(f) + S_2^{(I)}(f)$$

where the $S_j^{(I)}(f)$'s are nonnegative, nondecreasing functions with $S_j^{(I)}(-1/2) = 0$ and are of the following types:

- $S_1^{(I)}(\cdot)$ is absolutely continuous, i.e., its derivative exists for almost all f and is equal almost everywhere to an sdf $S(\cdot)$ such that

$$S^{(I)}(f) = \int_{-1/2}^{f} S(f')df'.$$

- $S_2^{(I)}(\cdot)$ is a step function with jumps of size $\{p_l\} : \ l = 1, 2, \ldots\}$ at the points $\{f_l : \ l = 1, 2, \ldots\}$.

We consider the integrated spectrum to be a combination of two 'pure' forms :

(a) $S_1^{(I)}(f) \geq 0; S_2^{(I)}(f) = 0$. $\{X_t\}$ is said to have a **purely continuous spectrum** and $S(f)$ is absolutely integrable, with

$$\int_{-1/2}^{1/2} S(f)\cos(2\pi f\tau)\,df \quad \text{and} \quad \int_{-1/2}^{1/2} S(f)\sin(2\pi f\tau) \to 0,$$

as $\tau \to \infty$. But,

$$s_\tau = \int_{-1/2}^{1/2} e^{i2\pi f\tau} S(f)\,df = \int_{-1/2}^{1/2} S(f)\cos(2\pi f\tau)\,df + i\int_{-1/2}^{1/2} S(f)\sin(2\pi f\tau)\,df$$

Hence $s_\tau \to 0$ as $|\tau| \to \infty$. In other words, the acvs diminishes to zero (called "mixing condition").

(b) $S_1^{(I)}(f) = 0; S_2^{(I)}(f) \geq 0$.
Here the integrated spectrum consists entirely of a step function, and the $\{X_t\}$ is said to have a purely discrete spectrum or a line spectrum. The acvs for a process with a line spectrum never damps down to 0.

Examples see Figs. 18. and 19.

(a) white noise, ARMA process.

(b) harmonic process.

Note that other combinations are possible:

Example 2.1.1 White noise spectrum
Recall that a white noise process $\{\epsilon_t\}$ has acvs:

$$s_\tau = \begin{cases} \sigma_\epsilon^2 & \tau = 0 \\ 0 & \text{otherwise} \end{cases}$$

Therefore, the spectrum of a white noise process is given by:

$$S_\epsilon(f) = \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f\tau} = s_0 = \sigma_\epsilon^2.$$

i.e., white noise has a constant spectrum.

2.1.3 Spectral density function vs. autocovariance function

The sdf and acvs contain the same amount of information in that if we know one of them, we can calculate the other. However, they are often not equally informative.

- The sdf usually proves to be the more sensitive and interpretable diagnostic or exploratory tool.
- Figure 20 show the sdf and acvs for two different processes - one with two pseudo periodicities and one with three.
  - The sdf is able to distinguish between the processes while the acvs's are not noticeably different.
  - $\text{dB} = 10 \log_{10}(\text{power})]$.

2.2 SAMPLING AND ALIASING

So far we have only looked at discrete time series $\{X_t\}$. However, such a process is usually obtained by sampling a continuous time process at equal intervals Δt, i.e., for a sampling interval $\Delta t > 0$ and an arbitrary time offset t_0, we can define a discrete time process through

$$X_t \equiv X(t_0 + t\Delta t), \qquad t = 0, \pm 1, \pm 2, \ldots .$$

If $\{X(t)\}$ is a stationary process with, say, sdf $S_{X(t)}(\cdot)$ and acvf $s(\tau)$, then $\{X_t\}$ is also a stationary process with, say, sdf $S_{X_t}(\cdot)$ and acvs $\{s_\tau\}$.

It can be shown that when $S_{X(t)}^{(I)}$ is differentiable:

$$S_{X_t}(f) = \sum_{k=-\infty}^{\infty} S_{X(t)}\left(f + \frac{k}{\Delta t}\right) \quad \text{for} \quad |f| \leq \frac{1}{2\Delta t}.$$

Thus, the discrete time sdf at f is the sum of the continuous time sdf at frequencies $f \pm \frac{k}{\Delta t}, \quad k = 0, 1, 2, \ldots$.

The frequency $1/(2\Delta t)$ is called the Nyquist frequency; previously we have taken $\Delta t = 1$, so that the frequency range was $|f| \leq \frac{1}{2}$.

If $S_{X(t)}$ is essentially zero for $|f| > 1/(2\Delta t)$ we can expect good correspondence between $S_{X_t}(f)$ and $S_{X(t)}(f)$ for $|f| \leq 1/(2\Delta t)$ (since

$$S_{X(t)}(f \pm k/(2\Delta t)) \approx 0$$

for $k = 1, 2, \ldots$).

If $S_{X(t)}$ is large for some $|f| > 1/(2\Delta t)$, the correspondence can be quite poor, and an estimate of S_{X_t} will not tell us much about $S_{X(t)}$.

Figure 21 illustrates this idea.

2.3 LINEAR FILTERING

A **linear time invariant (LTI) filter** L that transforms an input sequence $\{x_t\}$ into an output sequence $\{y_t\}$ has the following three properties:

1. **Scale-preservation**:

$$L\{\{\alpha x_t\}\} = \alpha L\{\{x_t\}\}.$$

2. **Superposition:**

$$L\{\{x_{t,1} + x_{t,2}\}\} = L\{\{x_{t,1}\} + L\{\{x_{t,2}\}.$$

3. **Time invariance:** If

$$L\{\{x_t\}\} = \{y_t\}, \quad \text{then} \quad L\{\{x_{t+\tau}\}\} = \{y_{t+\tau}\}.$$

Where τ is integer-valued, and the notation $\{x_{t+\tau}\}$ refers to the sequence whose t-th element is $x_{t+\tau}$.

Suppose we use a sequence with t-th element $\exp(i2\pi ft)$ as the input to a LTI digital filter: Let $\xi_{f,t} = \{e^{i2\pi ft}\}$, and let $y_{f,t}$ denote the output function:

$$y_{f,t} = L\{\xi_{f,t}\}.$$

By properties [1] and [3]:

$$y_{f,t+\tau} = L\{\xi_{f,t+\tau}\} = L\{e^{i2\pi f\tau}\xi_{f,t}\} = e^{i2\pi f\tau}L\{\xi_{f,t}\} = e^{i2\pi f\tau}y_{f,t}.$$

In particular, for $t = 0$:

$$y_{f,\tau} = e^{i2\pi f\tau}y_{f,0}.$$

Now set $\tau = t$:

$$y_{f,t} = e^{i2\pi ft}y_{f,0}.$$

Thus, when $\xi_{f,t}$ is input to the LTI digital filter, the output is the same function multiplied by some constant, $y_{f,0}$, which is independent of time but will depend on f. Let $G(f) = y_{f,0}$. Then

$$L\{\xi_{f,t}\} = \xi_{f,t} G(f).$$

$G(f)$ is called the transfer function or frequency response function of L. We can write

$$G(f) = |G(f)| e^{i\theta(f)}$$

where,

$$\begin{array}{ll} |G(f)| & \text{gain} \\ \theta(f) = \arg\{G(f)\} & \text{phase} \end{array}$$

Any LTI digital filter can be expressed in the form:

$$L\{\{X_t\}\} = \sum_{u=-\infty}^{\infty} g_u X_{t-u} \equiv \{Y_t\},$$

where $\{g_u\}$ is a real-valued deterministic sequence called the impulse response sequence. Note,

$$L\{\{e^{i2\pi f t}\}\} = \sum_{u=-\infty}^{\infty} g_u e^{i2\pi f(t-u)} = e^{i2\pi f t} G(f),$$

with

$$G(f) = \sum_{u=-\infty}^{\infty} g_u e^{-i2\pi f u} \quad \text{for} \quad |f| \leq \frac{1}{2}.$$

Note:

$$\{g_u\} \longleftrightarrow G(f) \quad \text{(F.T. pair)}.$$

We have,

$$Y_t = \sum_u g_u X_{t-u}$$

Recall,

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi ft}\, dZ_X(f) \qquad Y_t = \int_{-1/2}^{1/2} e^{i2\pi ft}\, dZ_Y(f),$$

which implies that

$$\begin{aligned} \int e^{i2\pi ft}\, dZ_Y(f) &= \sum_u g_u \int_{-1/2}^{1/2} e^{i2\pi f(t-u)}\, dZ_X(f) \\ &= \int_{-1/2}^{1/2} e^{i2\pi ft} G(f)\, dZ_X(f) \end{aligned}$$

Therefore

$$dZ_Y(f) = G(f)\, dZ_X(f) \; ; \qquad (1:1)$$

and

$$E\{|dZ_Y(f)|^2\} = |G(f)|^2 E\{|dZ_X(f)|^2\},$$

and if the spectral densities exist

$$S_Y(f) = |G(f)|^2 S_X(f).$$

This relationship can be used to determine the sdf's of discrete parameter stationary processes.

2.4 SDFS BY LTI FILTERING

Example 2.4.1 q-th order moving average: MA(q),

$$X_t = \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q},$$

with usual assumptions (mean zero). Define

$$L\{\{\epsilon_t\}\} = \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q},$$

so that $\{X_t\} = L\{\{\epsilon_t\}\}$. To determine $G(f)$, input $e^{i2\pi ft}$:

$$\begin{aligned} L\left\{\{e^{i2\pi ft}\}\right\} &= e^{i2\pi ft} - \theta_{1,q}e^{i2\pi f(t-1)} - \ldots \theta_{q,q}e^{i2\pi f(t-q)} \\ &= e^{i2\pi ft}\left[1 - \theta_{1,q}e^{-i2\pi f} - \ldots - \theta_{q,q}e^{-i2\pi fq}\right], \end{aligned}$$

so that

$$G_\theta(f) = 1 - \theta_{1,q}e^{-i2\pi f} - \ldots - \theta_{q,q}e^{-i2\pi fq}.$$

Since,

$$S_X(f) = |G_\theta(f)|^2 S_\epsilon(f) \quad \text{and} \quad S_\epsilon(f) = \sigma_\epsilon^2,$$

we have

$$S_X(f) = \sigma_\epsilon^2 |1 - \theta_{1,q} e^{-i2\pi f} - \ldots - \theta_{q,q} e^{-i2\pi f q}|^2.$$

If we put $z = e^{-i\omega}$ where $\omega = 2\pi f$, then

$$G_\theta(z) = 1 - \theta_{1,q} z - \ldots - \theta_{q,q} z^q,$$

and

$$|G_\theta(f)|^2 = G_\theta(f) G_\theta^*(f) \equiv G_\theta(z) G_\theta(z^{-1})$$

But for invertibility, $G_\theta(z)$ has no roots inside or on the unit circle.

Since $|G_\theta(f)|^2$ treats $G_\theta(z)$ and $G_\theta(z^{-1})$ as equals, and the roots of $G_\theta(z)$ and $G_\theta(z^{-1})$ are inverses, it is not possible to tell whether a moving-average process is invertible from its spectrum.

This makes sense, since we cannot distinguish these cases using the acvs either.

Example 2.4.2 p-th order autoregressive process: AR(p),

$$X_t - \phi_{1,p} X_{t-1} - \ldots - \phi_{p,p} X_{t-p} = \epsilon_t$$

Define

$$L\{\{X_t\}\} = X_t - \phi_{1,p} X_{t-1} - \ldots - \phi_{p,p} X_{t-p},$$

so that $L\{\{X_t\}\} = \{\epsilon_t\}$. By analogy to MA($q$)

$$G_\phi(f) = 1 - \phi_{1,p} e^{-i2\pi f} - \ldots - \phi_{p,p} e^{-i2\pi fp}.$$

Since,

$$|G_\phi(f)|^2 S_X(f) = S_\epsilon(f) \quad \text{and} \quad S_\epsilon(f) = \sigma_\epsilon^2,$$

we have

$$S_X(f) = \frac{\sigma_\epsilon^2}{|1 - \phi_{1,p} e^{-i2\pi f} - \ldots - \phi_{p,p} e^{-i2\pi fp}|^2}$$

Interpretation of AR spectra
Recall that for an AR process we have characteristic equation

$$1 - \phi_{1,p} z - \phi_{2,p} z^2 - \ldots - \phi_{p,p} z^p$$

and the process is stationary if the roots of this equation lie outside the unit circle.

Example 2.4.3 Consider an AR(2) process with complex characteristic roots, these roots must form a complex conjugate pair:

$$z = \frac{1}{r} e^{-i2\pi f'}, \qquad z = \frac{1}{r} e^{i2\pi f'}$$

and we can write

$$\begin{aligned} 1 - \phi_{1,p} z - \phi_{2,p} z^2 &= (rz - e^{-i2\pi f'})(rz - e^{i2\pi f'}) = r^2 z^2 - zr(e^{-i2\pi f'} + e^{i2\pi f'}) + 1 \\ &= r^2 z^2 - 2zr\cos(2\pi f') + 1 \end{aligned}$$

and the AR process can be written

$$(r^2B^2 - 2r\cos(2\pi f')B + 1)X_t = \epsilon_t$$

$$\Rightarrow X_t = 2r\cos(2\pi f')X_{t-1} - r^2X_{t-2} + \epsilon_t$$

The spectrum can be written in terms of the complex roots, by substituting $z = e^{-i2\pi f}$ in the characteristic equation.

$$S_X(f) = \frac{\sigma_\epsilon^2}{|re^{-i2\pi f} - e^{-i2\pi f'}|^2|re^{-i2\pi f} - e^{i2\pi f'}|^2}$$

Now,

$$\begin{aligned}
|re^{-i2\pi f} - e^{-i2\pi f'}|^2 &= |e^{-i2\pi f}(r - e^{-i2\pi(f'-f)})|^2 \\
&= (r - e^{-i2\pi(f'-f)})(r - e^{i2\pi(f'-f)}) \\
&= r^2 - r(e^{-i2\pi(f'-f)} + e^{i2\pi(f'-f)}) + 1 \\
&= r^2 - 2r\cos(2\pi(f'-f)) + 1
\end{aligned}$$

similarly,

$$|re^{-i2\pi f} - e^{i2\pi f'}|^2 = r^2 - 2r\cos(2\pi(f' + f)) + 1$$

giving,

$$S_X(f) = \frac{\sigma_\epsilon^2}{(r^2 - 2r\cos(2\pi(f' + f)) + 1)(r^2 - 2r\cos(2\pi(f' - f) + 1)}$$

The spectrum will be at its largest when denominator is at its smallest - when r is close to 1 this occurs when $f \approx \pm f'$. Also notice that at $f = \pm f'$ as $r \to 1$ (from below as $0 < r < 1$) so the spectrum becomes larger.

Generally speaking complex roots will induce a peak in the spectrum, indicating a tendency towards a cycle at frequency f'. Also, the larger the value of r the more dominant the cycle. This may be termed *pseudo-cyclical* behaviour (recall that a deterministic cycle will show up at a sharp spike – i.e., a line spectrum).

Example 2.4.4 $(p,q)-$th order autoregressive, moving average process: ARMA(p,q),

$$X_t - \phi_{1,p}X_{t-1} - \ldots - \phi_{p,p}X_{t-p} = \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q}$$

If we write this as

$$X_t - \phi_{1,p}X_{t-1} - \ldots - \phi_{p,p}X_{t-p} = Y_t;$$

$$Y_t = \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q},$$

then we have

$$|G_\phi(f)|^2 S_X(f) = S_Y(f) \qquad \text{and} \qquad S_Y(f) = |G_\theta(f)|^2 S_\epsilon(f)$$

so that

$$S_X(f) = S_\epsilon(f)\frac{|G_\theta(f)|^2}{|G_\phi(f)|^2} = \sigma_\epsilon^2 \frac{|1-\theta_{1,q}e^{-i2\pi f} - \ldots - \theta_{q,q}e^{-i2\pi fq}|^2}{|1-\phi_{1,p}e^{-i2\pi f} - \ldots - \phi_{p,p}e^{-i2\pi fp}|^2}$$

Example 2.4.5 Differencing
Let $\{X_t\}$ be a stationary process with sdf $S_X(f)$. Let $Y_t = X_t - X_{t-1}$. Then

$$\begin{aligned} L\left\{\{e^{i2\pi ft}\}\right\} &= e^{i2\pi ft} - e^{i2\pi f(t-1)} \\ &= e^{i2\pi ft}(1 - e^{-i2\pi f}) \\ &= e^{i2\pi ft}G(f), \end{aligned}$$

so

$$\begin{aligned} |G(f)|^2 &= |1 - e^{-i2\pi f}|^2 = |e^{-i\pi f}(e^{i\pi f} - e^{-i\pi f}|^2 \\ &= |e^{-i\pi f}2i\sin(\pi f)|^2 = 4\sin^2(\pi f). \end{aligned}$$