2 Descriptive methods of analysis

There is a large literature devoted to deriving methods for estimating and removing, or adjusting, various features of a time series. But in the first instance, it is important to look at your data, and to use simpler methods of analysis where appropriate.

- Time plots.
- Smoothing: looking for trends, seasonal effects
  - moving averages
  - regression models
  - spline smoothing.
- Differencing: removal of trend and seasonal effects.
- Autocovariance and autocorrelation functions.
- Variogram.
- Periodogram.

2.1 Time Plots

- Always start by plotting the series in *time order*. This is the most basic graphical representation; examine the main (qualitative) features of the graph.

  What might these be?

- Options are point or line plots; the latter usually enhance the display especially if you want to compare several series on the same plot. However, joining consecutive points can give a false sense of continuous observation, and any gaps or jumps may not be so obvious.

- **Aspect ratio**: relative sizes of the y and x axes.

Although graphical representations are assumed 'trivial' or easy, good representations require thought. There are now several books on the subject.

Example: Lynx data. Figure 2.1 shows two time plots of a well-known dataset on the annual numbers of lynx trapped in the Mackenzie river district of North-West Canada from 1821 to 1934.

The data show a strong, approximate, 10 year cycle and perhaps also a 50 to 60 year cycle. In the second plot, we can clearly see the differences between the heights of successive peaks.

In the top plot, the y-axis is compressed relative to the x-axis, and the effect of this choice of aspect ratio is to highlight the asymmetry of the rise and fall within each cycle.

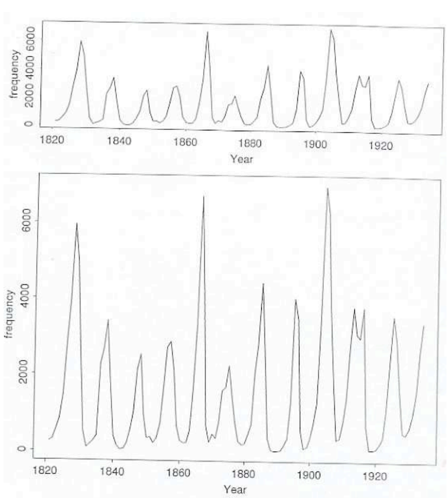

Fig. 2.1: Annual numbers of lynx trapped in the Mackenzie River District, Canada 1821–1934.

2.2 Smoothing

- Fundamentally an exploratory tool to
  - gain insight with respect to trends and/or seasonal effects, or
  - to remove trends and/or seasonal effects.
- **Example**: Monthly numbers of female deaths in the UK attributed to bronchitis, emphysema and asthma over the years 1974 to 1979; see Figure 2.2.

  Suppose we 'smooth' the data as follows. Let

  $$s_t = (y_{t-1} + y_t + y_{t+1})/3$$

  and instead of plotting y_t versus t, plot s_t versus t. Then join s_t by line segments and obtain the 'continuous' trace superimposed on the data. The smooth emphasizes the major features of the data, especially the seasonal pattern, but de-emphasizes the random fluctuations evident in the raw data.

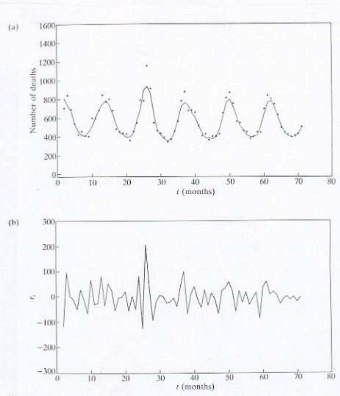

Fig. 2.3. Monthly returns of female deaths in the United Kingdom attributed to bronchitis, emphysema, and asthma over the years 1974 to 1979. (a) Raw data, y_t, shown as asterisks, and smoothed data, $s_t = (y_{t-1} + y_t + y_{t+1})/3$, shown as connected line segments. (b) Residuals, $r_t = y_t - s_t$.

Fig. 2.2: Female deaths in the UK from bronchitis, emphysema and asthma 1974–1979. From Diggle (1990).

- Smoothing: a decomposition of a time series y_t into a 'smooth' component s_t and a 'rough' component r_t:

$$y_t = s_t + r_t$$

(Note that this is *not* a model.)

Although it is tempting to equate the above decomposition to

$$Y(t) = \mu(t) + U(t),$$

suggesting the s_t is an estimate of $\mu(t)$, one needs more formal models and hypotheses.

Moving averages

- **Definition** A moving average of order $2p+1$ of a time series $\{y_t : \; t = 1, 2, \ldots, n\}$ is a time series defined by

$$s_t = \sum_{j=-p}^{p} w_j y_{t+j}, \qquad t = p+1, \ldots, n-p,$$

where p is a positive integer, the weights w_j are typically positive and satisfy $\sum_{j=-p}^{p} w_j = 1$. Usually $w_{-j} = w_j$, and $2p+1$ is called the *order* of the moving average.

- The definition leaves s_t undefined at the ends of the series. One approach is to sum from $j = \max(-p, 1-t)$ to $\min(p, n-t)$ and divide by the sum of weights included.

- Moving averages of odd order — to preserve the correspondence between y_t and s_t.

- **Seasonal effects**: handling these depends on whether the purpose is to

  - measure, or

  - to eliminate the seasonal effects

- Example: monthly deaths from lung diseases data.

  - To measure the effects for series with little or no trend, it is usually adequate to calculate the average for each month and compare these with the overall mean, either as a difference or a ratio. An alternative is to form a low-order moving average.

  - One way to eliminate the seasonal effects is to find a 13-point moving average with end weights $1/24$ and the rest $1/12$.

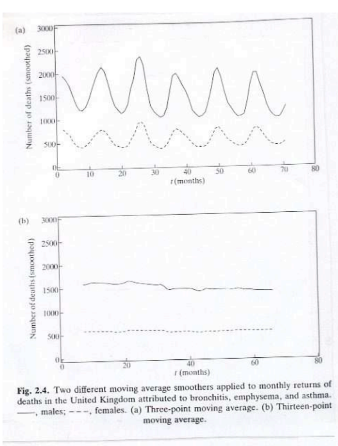

Fig. 2.3: Three-point (top) and 13-point (bottom) moving averages fitted to UK male and female deaths from bronchitis, emphysema and asthma 1974–1979. From Diggle (1990).

General comments:

- *How to choose an appropriate moving average?* Sometimes this is suggested by the context of the problem, or simply chosen for convenience.

- The example on Slide 2.5 with all w_j the same is called a *simple moving average of order 3*.

- Moving average smoothing is easy to implement and interpret (don't have to assume a parametric model).

- The method of repeated application of 3-point moving averages, inspecting the intermediate results, is a convenient (subjective) interactive way to proceed.

Regression models or curve fitting

- From a statistical point of view, fitting a regression model in time to the time series would appear to be a sensible way in which to model a trend.

- For example, polynomials are often used, although it is recognised that they are global models and as such rather inflexible. Thus we might take

$$s(t) = \sum_{j=0}^{p} \beta_j t^j$$

so that

$$\mathbf{y} = X\beta + \mathbf{e}$$

where X has (i, j)th element t_i^j and β is the vector of the β_j. This is a linear model, and assuming the 'errors' are independent and identically distributed, we have the least squares estimate of β,

$$\widehat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$$

- Low values of p (say 1 or 2) are useful for removing trends from time series, but large values are not usually useful.
- Copes with unequally spaced data.
- Least squares analysis straightforward.
- However, polynomial regression models of this form impose global assumptions about the nature of the data which are seldom warranted, and can produce artefacts.

Smoothing Spline

- To estimate the trend $\mu(t)$, we might consider least squares

$$\min_{\mu(t)} \sum_{i=1}^{n} (y_i - \mu(t_i))^2$$

to choose an arbitrary function $\mu(t)$. Obviously taking $\widehat{\mu}(t_i) = y_i$ will provide a perfect fit but give no indication of trend!

- Instead we consider minimizing

$$\sum_{i=1}^{n} (y_i - \mu(t_i))^2 + \alpha \int_{-\infty}^{\infty} \{\mu''(t)\}^2 dt$$

  - The second term is a penalty function, with α governing the amount of roughness we are prepared to tolerate.
  - If $\alpha = 0$, we reproduce the original data, that is, the fit is very ‘rough’. As α increases, the second term dominates, making the fitted values smooth.

- The ‘smoothest’ curve under this scheme is the straight line.

- The solution to the above problem is a function which satisfies the following:

  - $\widehat{\mu}(t)$ has a continuous first derivative everywhere

  - $\widehat{\mu}(t)$ is linear for $t < t_1$ and $t > t_n$

  - $\widehat{\mu}(t)$ is a cubic polynomial in t between each successive pair of t_i.

  - It can be shown that

$$\widehat{\mu}(t_i) = \sum_{j=1}^{n} w_{ij} y_j$$

    for certain weights w_{ij}, which is closely related to a moving average smoother.

- There is an automated way to select α, called *cross validation*. Unfortunately, for correlated data such as time series, the choice may not be very good.

- Spline smoothing can be advantageous as it can cope with arbitrary patterns of missing values in the data.

Example: *Mauna Loa Carbon Dioxide Concentration.*

The time series in Figure 2.4 represents monthly CO_2 concentrations in ppm (parts per million) from January 1959 to December 1990.

There is a strong yearly seasonal effect with CO_2 up in winter. There is also an obvious trend of increasing CO_2 concentration.

1. *Moving averages*: wiggly line is three-point moving average; smooth line is 13-point moving average.

2. *Polynomial regressions*: $p = 1$, linear regression; and $p = 2$, quadratic regression.

3. *Smoothing B-splines*: large α smooths the time series; α (approx. 10^{-10}) gives the wiggly line.

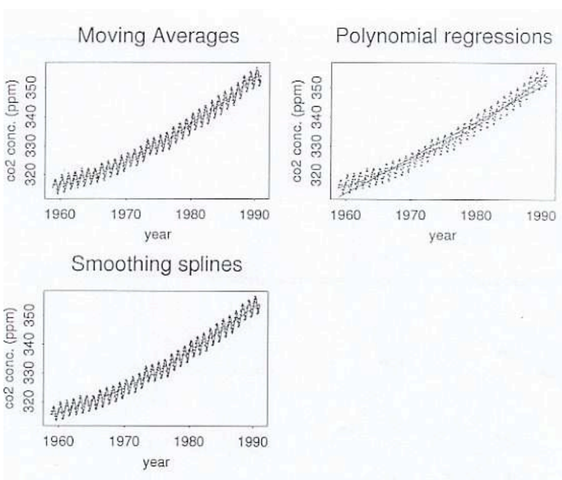

Fig. 2.4: Mauna Loa Carbon Dioxide Concentration: monthly concentrations in ppm from 1959 to 1990.

2.3 Differencing

The classical approach to time series is to remove trends, any seasonal effects, etc, with a view to emphasizing the ‘roughness’ or other features of the residuals.

- Simple approach to removing rather than highlighting trends.
- *First Difference*: $Dy_t = y_t - y_{t-1}$.
- *Higher-order Differences*: are obtained by repeated application, so for example, the second difference is

$$\begin{aligned} D^2 y_t = D(Dy_t) &= Dy_t - Dy_{t-1} \\ &= y_t - 2y_{t-1} + y_{t-2}. \end{aligned}$$

- If y_t consists of a kth degree polynomial in t, taking kth order difference of the series will remove this trend.

- If, in addition, the random component is stationary, $\{D^k y_t\}$ is also stationary.
- More importantly, it is sometimes the case that while y_t is not stationary, the differenced series $\{D^k y_t\}$ is stationary.
- If differencing is used to remove trend, the usual approach is to try Dy_t, $D^2 y_t$ and so on and examine the plots of these series, choosing the differencing that suggests stationarity.
- In practice, $k = 1$ or $k = 2$ is usually enough, and correspond to subtraction of simple moving averages of order 2 and 3.

- If an *increase* in variability is observed, this usually indicates that the series has been 'over-differenced'.

- **Seasonal effects**: can be eliminated by the use of appropriate differencing. For example, for monthly data, defining $B^k y_t = y_{t-k}$ to be the *backward shift operator* of order k, we take

$$z_t = y_t - B^{12} y_t = y_t - y_{t-12}$$

  We'll be using B with ARMA series later. Note that $Dy_t = y_t - By_t = (1 - B)y_t$.

Example: *Mauna Loa CO_2 concentration.* See Figure 2.5.

The dataset is available in R as `data(co2)`.

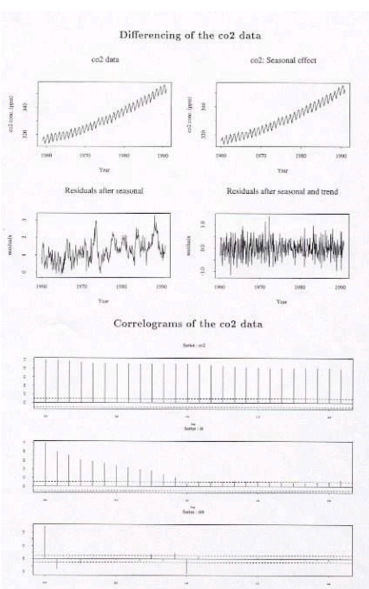

Fig. 2.5:

2.4 Autocovariance and autocorrelation functions

- Let $Y(t)$ be a stationary random function.
- The *autocovariance function of* $Y(t)$ is
$$\gamma(k) = \text{cov}\{Y(t), Y(t-k)\}$$
where k is the lag.
- The *autocorrelation function of* $Y(t)$ is defined to be
$$\rho(k) = \gamma(k)/\gamma(0),$$
since $\text{Var}(Y(t)) = \gamma(0)$.
- Important properties of $\rho(k)$ are:
  - $\rho(k) = \rho(-k)$
  - $-1 \leq \rho(k) \leq 1$
  - if $Y(t)$ and $Y(t-k)$ are independent, $\rho(k) = 0$, but note the converse is not true.

For the **discrete case**, $\{Y_t\}$ is a stationary random sequence and k is an integer, and we often write

- $\gamma_k =$ autocovariance coefficient of $\{Y_t\}$ and
- $\rho_k =$ autocorrelation coefficient of $\{Y_t\}$.

The acf is an important (albeit incomplete) summary of the *serial dependence* within a stationary random function or sequence.

Example: *A first-order autoregressive process*.

Consider the random sequence $\{Y_t\}$ defined by
$$Y_t = \alpha Y_{t-1} + Z_t, \quad (*)$$
where $\{Z_t\}$ is a sequence of white noise. When $\alpha = 1$, we obtain the simple random walk considered in previous examples. In Chapter 3, we will show that $\{Y_t\}$ is stationary if and only if $-1 < \alpha < 1$. Assume this for now.

We can show that $E(Y_t) = 0$, so the first condition for stationarity is satisfied. Multiplying both sides of (*) by Y_{t-k}, we can (eventually) show that

$$\gamma_k = \alpha^k \gamma_0$$

and hence that

$$\text{var}(Y_t) = \gamma_0 = \frac{\sigma^2}{1-\alpha^2}.$$

(We will work through this example in the lectures.)

Example: *Simulations with normally distributed white noise sequences $\{Z_t\}$ for different values of α.* See Figure 2.6 (which is Figure 2.9 from Diggle, 1990).

- $\alpha = -0.5$: the alternating pattern of autocorrelations for small lags reflects the tendency of the series $\{Y_t\}$ to zigzag about zero.
- $\alpha = 0.5$: the autocorrelations are strictly positive and this imparts some degree of smoothness in $\{Y_t\}$.
- $\alpha = 0.9$: the smoothness is now more evident; $\{Y_t\}$ takes relatively long excursions above and below zero (its average value).

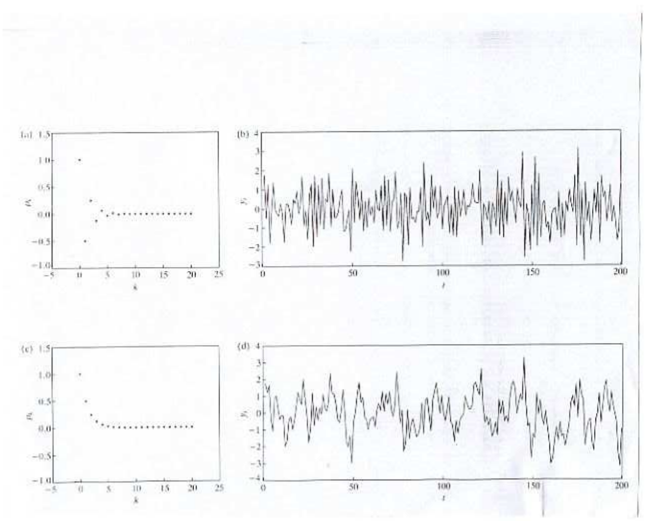

Fig. 2.6:

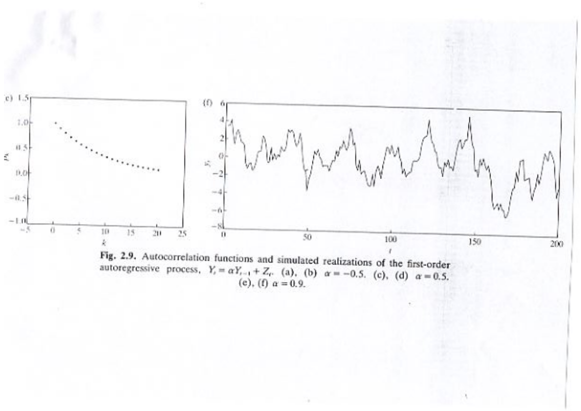

Fig. 2.9. Autocorrelation functions and simulated realizations of the first-order autoregressive process, $Y_t = \alpha Y_{t-1} + Z_t$. (a), (b) $\alpha = -0.5$. (c), (d) $\alpha = 0.5$. (e), (f) $\alpha = 0.9$.

Fig. 2.7: This is Figure 2.6 continued.

2.5 Estimating the autocorrelation function

Equally spaced series: the correlogram

- Let $\bar{y}$ be the sample mean.

- The kth sample autocovariance coefficient is

$$g_k = \frac{1}{n} \sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})$$

- The kth sample autocorrelation coefficient is

$$r_k = g_k / g_0,$$

  where g_0 is the sample variance.

- The *correlogram* is a plot of r_k against k.

- Bartlett (1946) showed that

$$r_k \sim N\left(0, \frac{1}{n}\right)$$

if process is white noise, so $|r_k| > 2/\sqrt{n}$ is significant at about the 5% level. However,

  - the more r_k that are examined the more likely a significant one will be found, even if the process is white noise.

  - the r_k are not independent;

  - Portmanteau test of white noise: under that null hypothesis,

$$Q_m = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}r_k^2 \sim \chi_m^2$$

  - A large value of Q_m suggests that the sample autocorrelations in the data are too large to be a sample from a white noise sequence.

  - The sensitivity of Q_m to various types of departure from white noise depends on the choice of m.

- The primary use of the correlogram is, however, to shed light on the nature of the serial dependence within a set of data. Thus we want to *relate* the form of the correlogram to various theoretical forms of the autocorrelations with a view to suggesting plausible models.

- The acf is valuable for determining the nature of the serial dependence in a time series.

- When $\{Y(t)\}$ jointly normal, ρ_k completely describes the process.

- In general though, the acf is an *incomplete* description of serial dependence in the sense that random processes whose realizations are qualitatively different can give rise to the same acf.

Some remarks:

1. We use the correlogram to shed light on the nature of the serial dependence within a time series. Thus the overall pattern is typically more important than the individual values of r_k.

2. One can relate the qualitative behaviour of the correlogram to various theoretical forms of the autocorrelation function, and this may suggest plausible models. The reliability of the correlogram for this purpose increases with the length of the series. Why?

3. Slow, approximate linear decay is typical of the behaviour of a correlogram of a non-stationary series whose theoretical autocorrelation function does not exist.

4. Trend and seasonality are usually detected by inspecting the graph of the (possibly transformed) series. However, they are also characterized by autocorrelation functions that are slowly decaying and nearly periodic respectively.

Unequally spaced series: the variogram

- Consider a stationary random function $Y(t)$.
- *Variogram*: $V(k) = \frac{1}{2}\mathsf{E}\left[\{Y(t) - Y(t-k)\}^2\right]$.
- Can show $V(k) = \gamma(0)(1 - \rho(k))$.
- For a time series $y(t_i):\ i = 1, 2, \ldots, n$,
  - calculate $v_{ij} = \frac{1}{2}\{y(t_i) - y(t_j)\}^2$ and $k_{ij} = t_i - t_j$ for all distinct $\frac{1}{2}n(n-1)$ pairs of observations.
  - *Empirical or sample variogram*: plot v_{ij} against k_{ij}. Improve the plot by averaging all values with common k_{ij}, call these values $\bar{v}(k)$.
- As k increases, usually $\rho(k)$ decays to zero. So the limiting value of $V(k)$ can be estimated by

$$v_\infty = g_0 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2$$

Hence

$$\hat{\rho}(k) = 1 - \bar{v}(k)/v_\infty$$

Comments

1. If $\bar{v}_k$ does not level off, then

- the serial dependence extends beyond the time span of the data, or

- the underlying process is non-stationary.

2. $V(k)$ exists for some non-stationary processes

e.g.

$$Y_t = Y_{t-1} + Z_t, \quad Y_0 = 0,$$

has $V(k) = \frac{1}{2}k\sigma^2$.

2.6 Impact of trend removal on autocorrelation structure

Recall our basic working model for a non-stationary random function:

$$Y(t) = \mu(t) + U(t).$$

If we knew $\mu(t)$ exactly, we could subtract it from the observed series $\{y(t_i)\}$, $i = 1, \ldots, n$, to obtain $\{u(t_i)\}$, a stationary series. In practice, $\mu(t)$ is usually not known and we estimate it. Then subtracting gives

$$r(t) = y(t) - \widehat{\mu}(t) \neq u(t).$$

That is, $\{r(t_i)\}$ is a 'corrupted' residual series.

One consequence of trend removal is that it induces spurious autocorrelations into the residual sequence.

Example: Suppose $\{U_t\}$ is a white noise sequence, and consider a moving average of order 3, so that

$$\begin{aligned} r_t &= u_t - (u_{t-1} + u_t + u_{t+1})/3 \\ &= (-u_{t-1} + 2u_t - u_{t+1})/3. \end{aligned}$$

Let $\{R_t\}$ with

$$R_t = (-U_{t-1} + 2U_t - U_{t+1})/3$$

be a stationary random sequence with autocovariance coefficient γ_k.

Then we can show that the *induced* autocorrelation coefficient of $\{R_t\}$ is

$$\rho_k = \begin{cases} 1 & \text{if } k = 0 \\ -\frac{2}{3} & \text{if } k = 1 \\ \frac{1}{6} & \text{if } k = 2 \\ 0 & \text{if } k = 3 \end{cases}$$

#

In general, the number of non-zero induced autocorrelations is one less than the order of the moving average.

For example, for a 3-point moving average, we get 2 induced autocorrelations.

But their magnitudes decrease as the order of the moving average decreases, e.g., using a $(2p+1)$-point (unweighted) moving average, the induced autocorrelation at lag 1 is

$$\rho_1 = -\frac{(p+1)}{p(2p+1)} \approx -\frac{1}{2p}$$

2.7 Periodogram

- Summary plot resulting from representing a time series as a sum of sinusoidal waves at various frequencies.

- Why? To elucidate the cyclic structure in the series, at frequencies not predictable in advance.

  In fact the periodogram is now more broadly viewed as a way of representing trends and other fluctuations in time series. It is especially useful for summarizing seasonality and short-term trends.

- Expressing the series in terms of its Fourier components forms the basis of *frequency domain methods*. This approach is especially important in engineering applications such as signal processing and structural design.

- A simple model:

$$y_t = \alpha \cos(\omega t) + \beta \sin(\omega t) + z_t$$

  where z_t is white noise, $t = 1, \dots, n$,

  - frequency $\omega = 2\pi/p$, $0 \leq \omega \leq \pi$
  - period p, assumed known,
  - amplitude $\sqrt{\alpha^2 + \beta^2}$
  - usually restrict to *Fourier frequencies*, $\omega_j = 2\pi j/n$, for some positive integer $j < n/2$, so that the period is n/j.

Treat as a linear model:

$$\boldsymbol{y} = X\boldsymbol{\theta} + \boldsymbol{z},$$

where $\boldsymbol{\theta} = (\alpha, \beta)^T$, and estimate α and β by least squares:

- $\hat{\boldsymbol{\theta}} = (X^T X)^{-1} X^T \boldsymbol{y}$

- $\hat{\alpha} = \frac{2}{n} \sum_{t=1}^{n} y_t \cos(\omega t)$

- $\hat{\beta} = \frac{2}{n} \sum_{t=1}^{n} y_t \sin(\omega t)$

- regression sum of squares:
$$\frac{n}{2}(\hat{\alpha}^2 + \hat{\beta}^2)$$
which has a χ_2^2 distribution under $H_0 : \alpha = \beta = 0$ (i.e., no cyclic effect), and assuming $\{Z_t\}$ normal. It is approximately χ_2^2 for large n and white noise.

- Generalize the simple model to one with several sinusoidal components: m is the largest integer less than $n/2$:
$$y_t = \sum_{k=1}^{m} \{\alpha_k \cos(\omega_k t) + \beta_k \sin(\omega_k t)\} + u_t$$
allows for more complex cyclic patterns.

- At the Fourier frequencies, $k > 0$, $\omega_k = 2\pi k/n$,
$$\hat{\alpha}_k = \frac{2}{n} \sum_{t=1}^{n} y_t \cos(\omega_k t)$$
$$\hat{\beta}_k = \frac{2}{n} \sum_{t=1}^{n} y_t \sin(\omega_k t)$$

- As m increases, we can achieve an orthogonal partitioning of progressively more of the variation in the series $\{y_t\}$ into sinusoidal components, each with 2 degrees of freedom.

- To achieve a complete partitioning of the sum of squares, add the extreme frequencies $\omega = 0$ and (if n is even), $\omega = \pi$.

- $\omega = 0$: sine terms vanish, cosines are identically 1. Thus model becomes $y_t = \alpha + z_t$. Clearly $\hat{\alpha} = \bar{y}$ and the associated contribution to the regression SS is $n\bar{y}^2$ on 1 df.

- $\omega = \pi$, n even:
$$\hat{\alpha}_{n/2} = \frac{1}{n}\sum_{t=1}^{n}(-1)^t y_t$$
and the associated SS on 1 df is
$$\frac{1}{n}\left\{\sum_{t=1}^{n}(-1)^t y_t\right\}^2.$$

- **Periodogram ordinates**
$$I(\omega) = \frac{1}{n}\left\{\left(\sum_{t=1}^{n} y_t \cos \omega t\right)^2 + \left(\sum_{t=1}^{n} y_t \sin \omega t\right)^2\right\}$$

- Thus if we include all the frequencies
$$y'y = I(0) + 2\sum_{j=1}^{m} I(2\pi j/n) + I(\pi),$$
the last term appearing only if n is even.

- **Periodogram**: plot of $I(\omega)$ vs ω, for $0 < \omega \leq \pi$.

- The periodogram is an attempt to separate out the various cyclic components via this decomposition into orthogonal parts.

- Exclude $\omega = 0$: little information on cyclic effects.

- Fourier frequencies: this is a restriction which provides simplified statistical properties because of the orthogonality induced. The decomposition will not work at other frequencies, and so can at best be an approximation if it is used.

Remarks:

The periodogram is a plot of the squared amplitudes times $4/n$ for various frequencies ω.

If the series contains a well-defined cyclic component then the periodogram can be expected to have a sharp peak at the appropriate value of ω.

Note though that in practice, such a peak is often masked because the variability in the $I(\omega)$ values can make the plot extremely irregular. Sometimes peaks appear even when there are no genuine cycles because one or more local maxima stand out relative to its neighbouring values.

The periodogram is the basic tool in estimating the *spectral density function* $f(\omega)$ (see Chapter 4).

Connection between the periodogram and the correlogram (see Chapter 4)

$$I(\omega) = g_0 + 2\sum_{k=1}^{n-1} g_k \cos(k\omega)$$

so that the periodogram is the discrete Fourier transform of the sample autocovariance function. Hence the *normalized periodogram*

$$I(\omega)/g_0 = 1 + 2\sum_{k=1}^{n-1} r_k \cos(k\omega)$$

is the discrete Fourier transform of the correlogram.

Thus mathematically, the two are equivalent. But from a practical view, the periodogram focuses on the cyclic nature of the data, and the correlogram focuses on the serial dependence.

Cumulative Periodogram

One application of the periodogram is that we can derive an alternative to the Box-Pierce Portmanteau statistic for testing the hypothesis that $\{Y_t\}$ is white noise.

For $j = 1, 2, \ldots, p$ where $p = [n/2]$, define

$$C_j = \sum_{k=1}^{j} I(2\pi k/n)$$

- Let $p\prime = p - 1$ and $U_j = \frac{C_j}{C_p}$.
- *Cumulative periodogram*: plot of U_j vs $j/p\prime$.
- *Under white noise*: all the 'true' amplitudes of the sinusoidal components should be 0; thus the periodogram ordinates should only differ because of sampling fluctuations and the cumulative periodogram should be linear from 0 to 1 as j runs from 1 to $p\prime$.

- Test statistic (Kolmogorov-Smirnov): largest vertical distance of the cumulative periodogram to the straight line.
- Approximate critical value: $\pm 1.358(\sqrt{p\prime} + 0.12 + 0.11/\sqrt{p\prime})$
- Plot includes two parallel straight lines implied by the 5% critical values for visual assessment of the test.

Discussion example: *Heroin purity and fatal heroin overdose*.

In a recent study, researchers in NSW used time series analysis to determine the role, if any, played by heroin purity in fatal heroin overdoses. A total of 322 heroin samples were analyzed in fortnightly periods between February 1993 and January 1995: the samples were taken from street seizures in south-western Sydney. Over the same period, a total of 61 overdose deaths occurred in the same region.

Figure 2.8 shows a lineplot of the heroin purity data, Figures 2.9 to 2.11 give the raw periodogram, the smoothed periodogram, and the cumulative periodogram, respectively.

We will analyse these data further in Computing Practical 3.

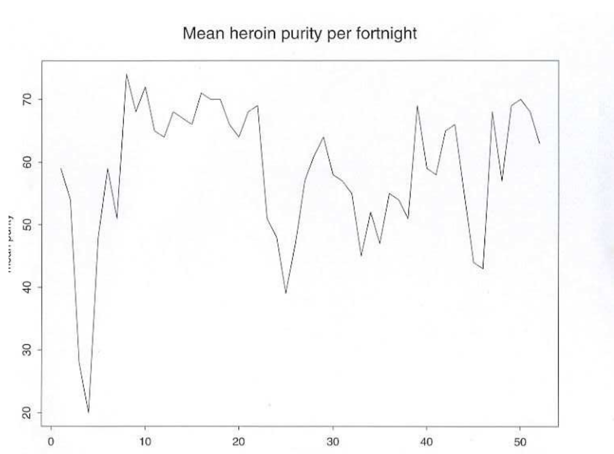

Fig. 2.8: Heroin samples taken from street seizures in south-western Sydney 1993–1995.

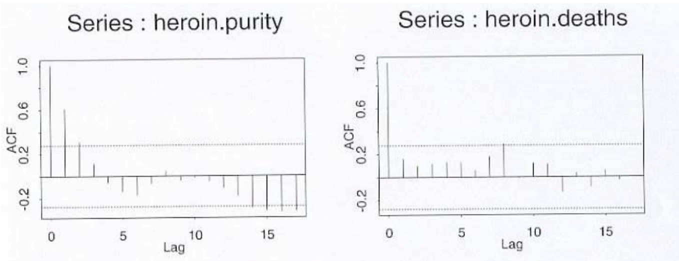

Fig. 2.9: Acfs for heroin samples and deaths from overdose, 1993–1995.

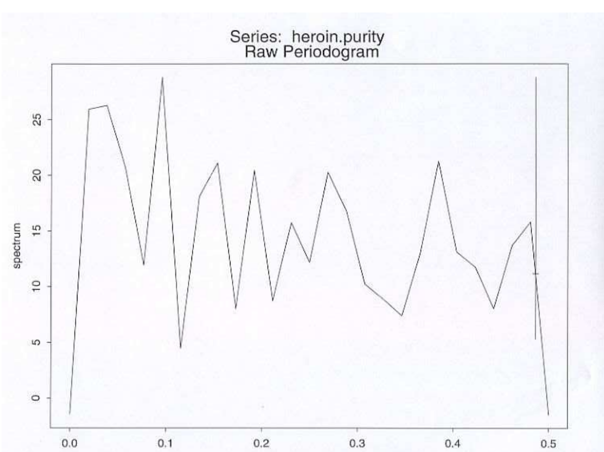

Fig. 2.10: The raw periodogram for heroin purity data.

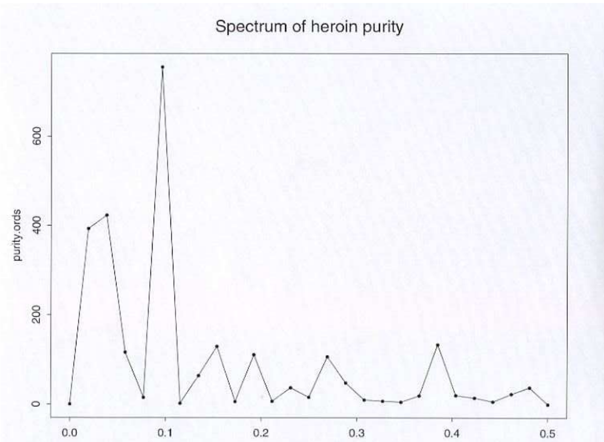

Fig. 2.11: The smoothed periodogram for heroin purity data.

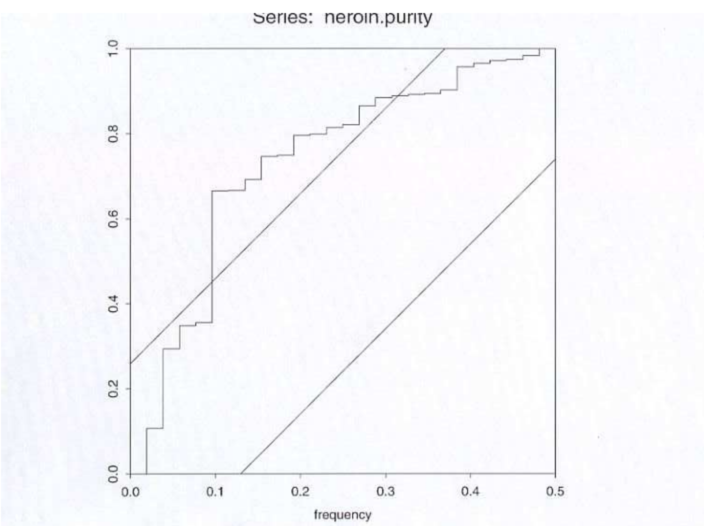

Fig. 2.12: The cumulative periodogram for heroin purity data.

2.8 A note on transformations

In general, there are three main reasons for making a transformation:

- *To stabilize the variance*: if there is an increasing trend, and the standard deviation is directly proportional to the mean, a log transformation is indicated.

- *To make seasonal effect additive*: if there is a trend in mean, and the size of the seasonal effect appears to increase with the mean, then it may be advisable to transform the data to make the seasonal effect constant from year to year. For example, if the seasonal effect is proportional to the mean, then the relationship is multiplicative and log transformation will induce additivity, although it only stabilizes the variance if the error term is also multiplicative.

- *To make data normally distributed*: model building and forecasting are usually carried out on the assumption that the data are normally distributed. This may not be the case, e.g., skewness leads to peaks (in one direction) in a time plot. Non-normal error distributions can be difficult to estimate and it may therefore be simpler to transform the data.

- A useful family of transformations is the Box-Cox family:

$$x_t = \begin{cases} (y_t^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ \log y_t & \text{if } \lambda = 0 \end{cases}$$

However, for forecasting from time series, studies have found little improvement in forecast performance following transformation. Transformation may correct some features of the data, but not others, and one often has to 'transform back' to provide useful interpretation, and this can induce bias. It is therefore better to avoid transformation wherever possible, except when the transformed variable has a direct physical interpretation.