Lecture 38

Cyclical Behavior and Periodicity: Part 2

Arnab Hazra

Frequency and period of a time series (recap)

- We measure frequency ω by cycles per time point.
- In the Johnson & Johnson data set, the predominant frequency of oscillation is one cycle per year (4 quarters), or $\omega = 0.25$ cycles per observation.
- The predominant frequency in the SOI and fish populations series is 1 cycle every 12 months, or $\omega = 0.083$ cycles per observation.
- For a discrete time series, we will need at least two points to determine a cycle, so the highest frequency, called the folding frequency, is 0.5.
- The period of a time series is the number of points in a cycle, i.e., $1/\omega$.
- The predominant period of the Johnson & Johnson series is $1/0.25$ or 4 quarters per cycle.
- The predominant period of the SOI series is 12 months per cycle.

Periodic process (recap)

- We first define a cycle as one complete period of a sine or cosine function defined over a unit time interval.
- We consider the periodic process

$$X_t = A\cos(2\pi\omega t + \phi)$$

  for $t = 0, \pm 1, \pm 2, \ldots$, where ω is a frequency index.
- Here A determines the height or amplitude of the function and ϕ, called the phase, determining the start point of the cosine function.
- We can introduce random variation in this time series by allowing the amplitude and phase to vary randomly.

Periodic process (contd., recap)

- For purposes of data analysis, it is easier to write X_t as

$$X_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t),$$

  where $U_1 = A\cos(\phi)$ and $U_2 = -A\sin(\phi)$.
- We then often take U_1 and U_2 to be normally distributed.
- The amplitude is $A = \sqrt{U_1^2 + U_2^2}$ and the phase is $\phi = \tan^{-1}(-U_2/U_1)$.
- Here, A and ϕ are independent random variables if U_1 and U_2 are independent standard normal random variables.
- Then $A^2 \sim \chi_2^2$ and $\phi \sim \mathrm{Unif}(-\pi, \pi)$.
- Straightforward Jacobian calculations show that the reverse is also true.

Moments of X_t (recap)

If we assume that U_1 and U_2 are uncorrelated random variables with mean 0 and variance σ^2, then

- Mean $E(X_t) = 0$.
- Covariance $\mathrm{Cov}(X_{t+h}, X_t) = \sigma^2 \cos(2\pi\omega h)$.
- Variance $\mathrm{Var}(X_t) = \sigma^2$
- If we observe U_1 and U_2, an estimate of σ^2 is the sample variance

$$S^2 = \frac{U_1^2 + U_2^2}{2-1} = U_1^2 + U_2^2.$$

Further generalization of periodic processes (recap)

- We can allow mixtures of periodic series with multiple frequencies and amplitudes,

$$X_t = \sum_{k=1}^{K} [U_{k1} \cos(2\pi\omega_k t) + U_{k2} \sin(2\pi\omega_k t)]$$

where U_{k1}, U_{k2}, for $k = 1, 2, \ldots, K$, are uncorrelated zero-mean random variables with variances σ_k^2, and the ω_k are distinct frequencies.
- Notice that the process is a sum of uncorrelated components, with variance σ_k^2 for frequency ω_k.
- The autocovariance function of X_t is

$$\mathrm{Cov}(X_{t+h}, X_t) = \sum_{k=1}^{K} \sigma_k^2 \cos(2\pi\omega_k h), \quad \mathrm{Var}(X_t) = \sum_{k=1}^{K} \sigma_k^2.$$

- Here $\widehat{\mathrm{Var}}(X_t) = \sum_{k=1}^{K} \hat{\sigma}_k^2 = \sum_{k=1}^{K} [U_{k1}^2 + U_{k2}^2]$.

Sum of periodic functions (recap)

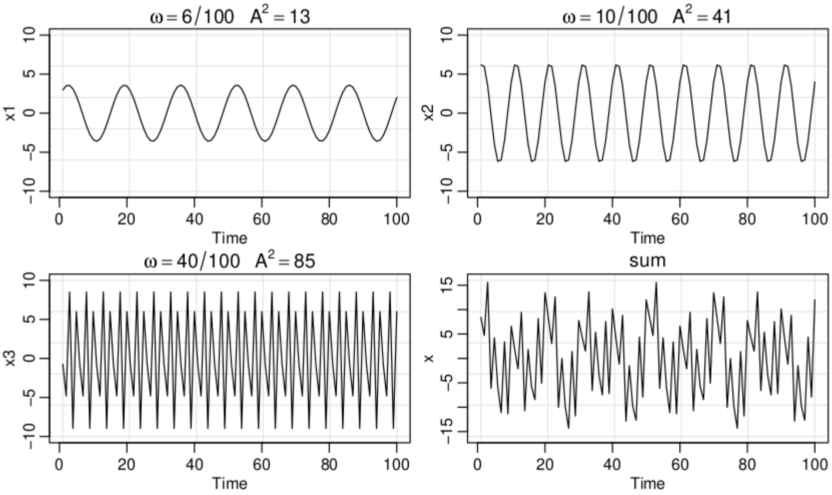

Fig. 4.1. *Periodic components and their sum as described in Example 4.1.*

Our goal (recap)

- The systematic sorting out of the essential frequency components, including their relative contributions, is one of the main objectives of *spectral analysis*.
- The moments of X_t only discuss the population properties but no statistical inference.
- If we can observe U_{k1} and U_{k2}, then $\hat{\sigma}_k^2 = U_{k1}^2 + U_{k2}^2$.
- In practice, we only observe $X_1, \ldots, X_T$ but not U_{k1}'s and U_{k2}'s.
- Hence, we next discuss the practical aspects of how, given data $X_1, \ldots, X_T$, to actually estimate the variance components $\sigma_k^2, k = 1, \ldots, K$.

Estimation: Periodogram

- Think of X_t as a color (waveform) made up of primary colors X_{t1}, X_{t2}, X_{t3} at various strengths (amplitudes).
- Then the periodogram is like a prism that decomposes the color X_t into its primary colors (spectrum). Hence the term spectral analysis.
- For any time series sample $X_1, \ldots, X_T$, where T is odd, we may write, exactly

$$X_t = a_0 + \sum_{j=1}^{(T-1)/2} [a_j \cos(2\pi t\, j/T) + b_j \sin(2\pi t\, j/T)],$$

  for $t = 1, \ldots, T$ and suitably chosen coefficients.
- If T is even, the representation can be modified by summing to $(T/2 - 1)$ and adding an additional component given by $a_{T/2} \cos(2\pi t \frac{1}{2}) = a_{T/2}(-1)^t$.

Estimation: Periodogram (contd.)

- Using the regression results, a_j and b_j are of the form

$$a_j = \frac{2}{T}\sum_{t=1}^{T} X_t \cos(2\pi t\, j/T), \quad b_j = \frac{2}{T}\sum_{t=1}^{T} X_t \sin(2\pi t\, j/T).$$

- We then define the scaled periodogram to be $P(j/T) = a_j^2 + b_j^2$.

- The scaled periodogram is simply the sample variance at each frequency component and consequently is an estimate of σ_j^2 corresponding to the sinusoid oscillating at a frequency of $\omega_j = j/T$.

Scaled periodogram

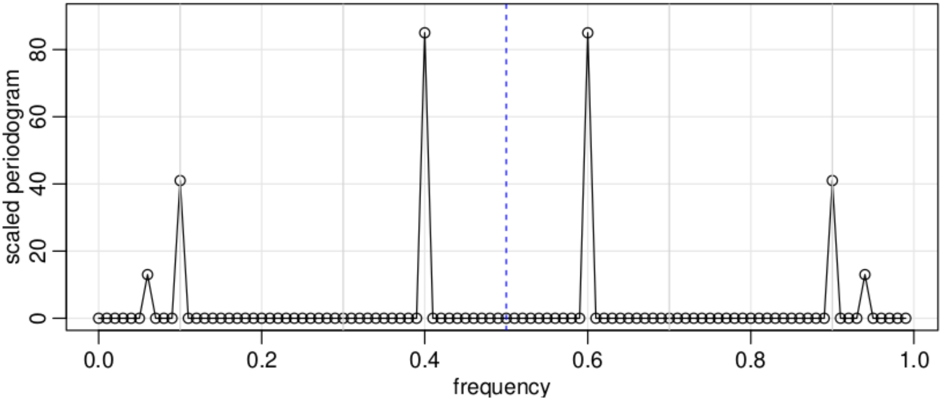

Fig. 4.2. *The scaled periodogram (4.12) of the data generated in Example 4.1.*

Estimation: Periodogram (contd.)

- Note that $P(j/T) = P(1 - j/T), j = 0, 1, \ldots, T-1$, so there is a mirroring effect at the folding frequency of 1/2.
- Consequently, the periodogram is typically not plotted for frequencies higher than the folding frequency.
- In addition, note that the heights of the scaled periodogram shown in the figure are
  $P\left(\frac{6}{100}\right) = P\left(\frac{94}{100}\right) = 13, \quad P\left(\frac{10}{100}\right) = P\left(\frac{90}{100}\right) = 41, \quad P\left(\frac{40}{100}\right) = P\left(\frac{60}{100}\right) = 85$
  and $P(j/T) = 0$ otherwise.

Periodogram for Star Magnitude

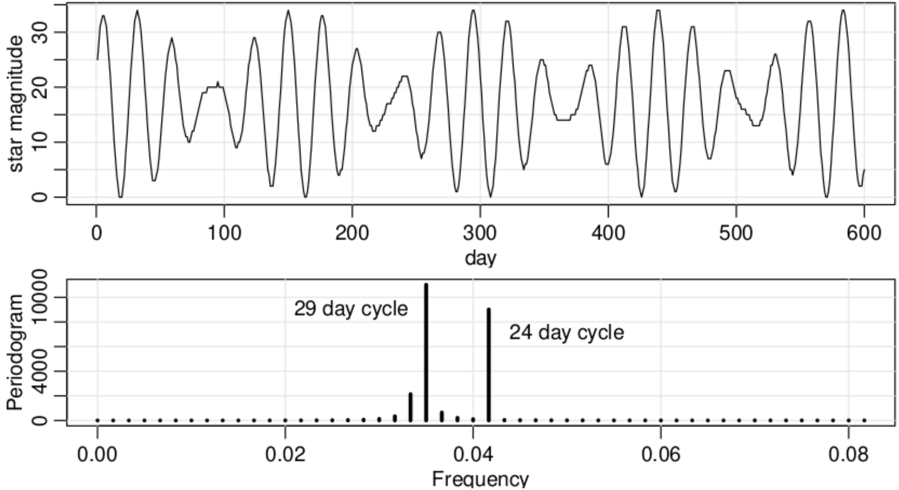

Fig. 4.3. *Star magnitudes and part of the corresponding periodogram.*

Star magnitude

- The data are the magnitude of a star taken at midnight for 600 consecutive days.
- Note that the 29($\approx 1/.035$) day cycle and the 24($\approx 1/.041$) day cycle are the most prominent periodic components of the data.
- We can interpret this result as we are observing an amplitude modulated signal.
- For example, suppose we are observing signal-plus-noise, $X_t = S_t + V_t$, where $S_t = \cos(2\pi\omega t)\cos(2\pi\delta t)$, and δ is very small.
- In this case, the process will oscillate at frequency ω, but the amplitude will be modulated by $\cos(2\pi\delta t)$.
- Since $2\cos(\alpha)\cos(\delta) = \cos(\alpha+\delta) + \cos(\alpha-\delta)$, the periodogram of data generated as X_t will have two peaks close to each other at $\alpha \pm \delta$.

Discrete Fourier transform

- It is not necessary to run a large regression to obtain the values of a_j and b_j because they can be computed quickly if T is a highly composite integer.
- The discrete Fourier transform (DFT) is a complex-valued weighted average of the data given by

$$d(j/T) = T^{-1/2} \sum_{t=1}^{T} X_t \exp[-2\pi\iota \; t \; j/T]$$

  for $j = 0, 1, \ldots, T-1$.
- Here the frequencies j/T are the Fourier or fundamental frequencies.
- Because of a large number of redundancies in the calculation, $d(j/T)$ may be computed quickly using the fast Fourier transform (FFT).

Periodogram

- Here
$$d(j/T) = T^{-1/2}\left\{\sum_{t=1}^{T} X_t \cos(2\pi t\, j/T) - \iota \sum_{t=1}^{T} X_t \sin(2\pi t\, j/T)\right\}.$$
- Periodogram is defined as
$$|d(j/T)|^2 = T^{-1}\left(\sum_{t=1}^{T} X_t \cos(2\pi t\, j/T)\right)^2 + T^{-1}\left(\sum_{t=1}^{T} X_t \sin(2\pi t\, j/T)\right)^2.$$
- Recall that the scaled periodogram is
$$P(j/T) = \left(\frac{2}{T}\sum_{t=1}^{T} X_t \cos(2\pi t\, j/T)\right)^2 + \left(\frac{2}{T}\sum_{t=1}^{T} X_t \sin(2\pi t\, j/T)\right)^2.$$
- Hence, $P(j/T) = 4T^{-1}|d(j/T)|^2$.

Thank you!