LECTURE EIGHT

Recap from last lecture

$$y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + \varepsilon_t$$

$t = 1, \dots, n$ $\qquad$ f: main parameters $\qquad$ $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$

$y_1, \dots, y_n$

$$RSS(f) = \min_{\beta_0, \beta_1, \beta_2} \sum_{t=1}^{n} \left[y_t - \beta_0 - \beta_1 \cos 2\pi f t - \beta_2 \sin 2\pi f t\right]^2$$

We can restrict f to $[0, 0.5]$

- Take a grid of f-values.
- Compute $RSS(f)$
- Minimize to obtain $\hat{f}$.

For efficient computation, we need a more explicit expression for $RSS(f)$.

Fourier Frequencies: f $\quad [0, 0.5]$

(nf must be an integer)

$$\begin{cases} \left\{0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \dots, \frac{n-1}{2n}\right\} & \text{if } n \text{ is odd} \\ \left\{0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \dots, \frac{n/2}{n}\right\} & \text{if } n \text{ is even} \end{cases}$$

LAST CLASS, we derived

$RSS(f) =$

when f is a Fourier frequency & $0 < f < \frac{1}{2}$

$$\sum_{t=1}^{n}(y_t-\bar{y})^2 - \frac{2}{n}\left(\sum_{t=1}^{n} y_t \cos 2\pi f t\right)^2 - \frac{2}{n}\left(\sum_{t=1}^{n} y_t \sin 2\pi f t\right)^2$$

$$\boxed{\begin{aligned} y\cdot\cos &= \sum_{t=1}^{n} y_t \cos 2\pi f t \\ y\cdot\sin &= \sum_{t=1}^{n} y_t \sin 2\pi f t \end{aligned}}$$

$$RSS(f) = \sum_{t=1}^{n}(y_t-\bar{y})^2 - \frac{2}{n}\left[(y\cdot\cos)^2 + (y\cdot\sin)^2\right]$$

$$I(f) = \frac{1}{n}\left\{(y\cdot\cos)^2 + (y\cdot\sin)^2\right\}$$

$0 < f < \frac{1}{2}$

f is a Fourier frequence

$$\boxed{RSS(f) = \sum_{t=1}^{n}(y_t-\bar{y})^2 - 2I(f)}$$

Periodogram

Minimizing $RSS(f)$ over Fourier frequencies $\Longleftrightarrow$ Maximizing $I(f)$ over Fourier frequencies

$$I(f) = \frac{1}{n}\left[(y\cdot\cos)^2 + (y\cdot\sin)^2\right]$$

$\left[e^{i\theta} = \cos\theta + i\sin\theta\right]$

$$\sum_{t=1}^{n} y_t e^{-2\pi i f t} \doteq y\cdot e^{2\pi i f t}$$

$$I(f) = \frac{1}{n} \left| \underbrace{y \cdot e^{2\pi i f t}} \right|^2$$

Real part: $y \cdot \cos$

Imaginary part: $-y \cdot \sin$

Dot product between complex vectors:

$a = (a_1 \cdots a_n)$ $\quad b = (b_1 \cdots b_n)$

$$a \cdot b = \langle a, b \rangle = \sum_{j=1}^{n} a_j \overline{b_j} \rightarrow \text{complex conjugate of } b_j$$

$$y \cdot e^{2\pi i f t} = \sum_{t=1}^{n} y_t e^{-2\pi i f t}$$

Periodogram: $\underbrace{I(f)} = \frac{1}{n} \left| \underbrace{y \cdot e^{2\pi i f t}} \right|^2$]

DFT (Discrete Fourier Transform)

Data $y_0, y_1 \cdots y_{n-1}$ (often represented by the vector y)

DFT:

$$b_j = \underbrace{\sum_{t=0}^{n-1} y_t \exp\left(-2\pi i \frac{j}{n} t\right)}_{y \cdot \exp\left(\underbrace{2\pi i \frac{j}{n} t}\right)}$$

b_j represents the Fourier frequency $\frac{j}{n}$

$$y = (y_0 \cdots y_{n-1})$$

$$u^j = \left(\exp\left(2\pi i \tfrac{j}{n} t\right), t = 0, 1, \ldots, n-1\right)$$

↳ Data coming from the complex sinusoid of frequency $\frac{j}{n}$.

$$b_j = y \cdot u^j = \langle y, u^j \rangle$$

Indexing starting at 0 vs 1

$$b_j = \sum_{t=0}^{n-1} y_t \exp\left(-2\pi i \tfrac{j}{n} t\right)$$

$$\tilde{b}_j = \sum_{t=1}^{n} y_t \exp\left(-2\pi i \tfrac{j}{n} t\right)$$

$$b_j = \text{first data point} \times e^{-2\pi i \frac{j}{n}(0)} + \text{2nd data point } e^{-2\pi i \frac{j}{n}(1)} + \cdots + t^{\text{th}} \text{ data point } e^{-2\pi i \frac{j}{n}(t-1)} + \cdots$$

$$\tilde{b}_j = \text{first data point} \times e^{-2\pi i \frac{j}{n}(1)} + \text{2nd data point } e^{-2\pi i \frac{j}{n}(2)} + \cdots + t^{\text{th}} \text{ data point } e^{-2\pi i \frac{j}{n} t} + \cdots$$

$$b_j = \tilde{b}_j \exp\left(\frac{2\pi i j}{n}\right)$$

Check then $|b_j| = |\tilde{b}_j|$

$$\boxed{b_j = \sum_{t=0}^{n-1} y_t \exp\left(-2\pi i \frac{j}{n} t\right) \quad j = 0, 1, \ldots, n-1}$$

$(b_0, \ldots, b_{n-1})$ is called the DFT of $y_0, \ldots, y_{n-1}$

Note: Data $y_0 \ldots y_{n-1}$ are always real but the DFT b_j can be complex

$$I\left(\frac{j}{n}\right) = \frac{|b_j|^2}{n} \quad \text{for } 0 < \frac{j}{n} < \frac{1}{2}$$

$$RSS\left(\frac{j}{n}\right) = \sum (y_t - \bar{y})^2 - 2\, I\left(\frac{j}{n}\right)$$

IMPORTANT: DFT b_j, $j = 0, 1, \ldots, n-1$ can be computed efficiently by the FFT algorithm. COOLEY & TUKEY.

FAST FOURIER TRANSFORM

$O(n \log n)$

$n = 4$ $n = 2$

More on DFT

$y_0, y_1, \dots y_{n-1}$

$$b_j = \sum_{t=0}^{n-1} y_t \exp\left(-2\pi i \frac{j}{n} t\right), \quad j = 0, 1, \dots, n-1$$

① $b_0 = \sum_{t=0}^{n-1} y_t$ } uninteresting

always real

② $$b_j = \underbrace{\sum_{t=0}^{n-1} y_t \cos 2\pi \frac{j}{n} t}_{Re(b_j)} - i \underbrace{\sum_{t=0}^{n-1} y_t \sin 2\pi \frac{j}{n} t}_{Im(b_j)}$$

Generally b_j will be complex.

(Sometimes b_j can be real e.g n even $j = \frac{n}{2}$)

③ $$b_{n-j} = \sum_{t=0}^{n-1} y_t \exp\left(-2\pi i \frac{n-j}{n} t\right)$$

$$= \sum_{t=0}^{n-1} y_t \exp\underbrace{\left(-2\pi i t\right)} \exp\left(2\pi i \frac{j}{n} t\right)$$

$$= \sum_{t=0}^{n-1} y_t \exp\left(2\pi i \frac{j}{n} t\right)$$

$$= \overline{\sum_{t=0}^{n-1} y_t \exp\left(-2\pi i \frac{j}{n} t\right)}$$

$e^{i\theta} = \cos\theta + i\sin\theta$

$e^{-i\theta} = \cos\theta - i\sin\theta$

$e^{i\theta} = \overline{e^{-i\theta}}$

$$= \sum_{t=0}^{n-1} y_t \exp\left(-2\pi i \frac{j}{n} t\right)$$ because y_t is real

$$= \bar{b}_j$$

Thus

$$\boxed{b_{n-j} = \bar{b}_j}$$

$n=10$
$j=5$
$b_5 = \bar{b}_5$

$n=11$ $\quad y_0, y_1, \ldots, y_{10}$

$$b_0, \underbrace{b_1, b_2, b_3, b_4, b_5}_{\text{complex}}, \underbrace{\bar{b}_5, \bar{b}_4, \bar{b}_3, \bar{b}_2, \bar{b}_1}$$

b_0 real

$$I\left(\frac{j}{n}\right) = \frac{|b_j|^2}{n}, \qquad I\left(\frac{6}{11}\right) = I\left(\frac{5}{11}\right)$$

$n=10$

$$b_0, b_1, b_2, b_3, b_4, b_5, \bar{b}_4, \bar{b}_3, \bar{b}_2, \bar{b}_1$$

b_5 real

③ You can recover the data from the DFT.

$$b_0, b_1, b_2, \ldots, b_{n-1}$$

$$y_t = \frac{1}{n} \sum_{j=0}^{n-1} b_j \exp\left(2\pi i \frac{j}{n} t\right) \rightarrow \text{Inverse DFT}$$

$$b_j = \sum_{t=0}^{n-1} y_t \exp\left(-2\pi i \frac{j}{n} t\right) \rightarrow \text{Defn of DFT}$$

Proof of IDFT

We use orthogonality of complex sinusoidal vectors.

$$u^j = \left(\exp\left(2\pi i \frac{j}{n} t\right),\ t = 0,1,\dots,n-1\right) \quad j = 0,1,\dots,n-1$$

$$= \left(1, \exp\left(2\pi i \frac{j}{n}(1)\right), \exp\left(2\pi i \frac{j}{n}(2)\right), \dots, \exp\left(2\pi i \frac{j}{n}(n-1)\right)\right)$$

$b_j = \langle y, u^j \rangle \rightarrow$ Defn of DFT

FACT: $u^0, u^1, \dots, u^{n-1}$ satisfy:

$$\underbrace{\langle u^j, u^k \rangle}_{\boxed{\exp(2\pi i f t)}} = \begin{cases} 0 & \text{if } j \neq k \\ n & \text{if } j = k \end{cases}$$

$u^0, u^1, \dots, u^{n-1} \rightarrow$ vectors of length n
orthogonal
& constant length

$\Downarrow$

they form a basis for all vectors of length n.

$\Downarrow$

every vector can be written as a linear combination of $u^0, u^1, \dots, u^{n-1}$

$$y = a_0 u^0 + a_1 u^1 + \cdots + a_{n-1} u^{n-1}$$

Take inner product on both sides with u^j.

$$\langle y, u^j \rangle = a_0 \langle u^0, u^j \rangle + a_1 \langle u^1, u^j \rangle + \cdots + a_{n-1} \langle u^{n-1}, u^j \rangle$$

$$= a_j \langle u^j, u^j \rangle = a_j n$$

$$\Rightarrow a_j = \frac{1}{n} \langle y, u^j \rangle = \frac{b_j}{n}$$

We proved:

$$y = \frac{b_0}{n} u^0 + \frac{b_1}{n} u^1 + \cdots + \frac{b_{n-1}}{n} u^{n-1}$$

$$\Rightarrow y_t = \frac{b_0}{n} u_t^0 + \frac{b_1}{n} u_t^1 + \cdots + \frac{b_{n-1}}{n} u_t^{n-1}$$

$$y_t = \frac{1}{n} \sum_{j=0}^{n-1} b_j \exp\left(2\pi i \frac{j}{n} t\right) \rightarrow \text{Inverse DFT}$$

Data $\longrightarrow$ DFT

TIME DOMAIN ANALYSIS y_t

Modeling b_j

FREQUENCY DOMAIN ANALYSIS