Empirical likelihood method in statistical inference 3

Hiroaki Ogata

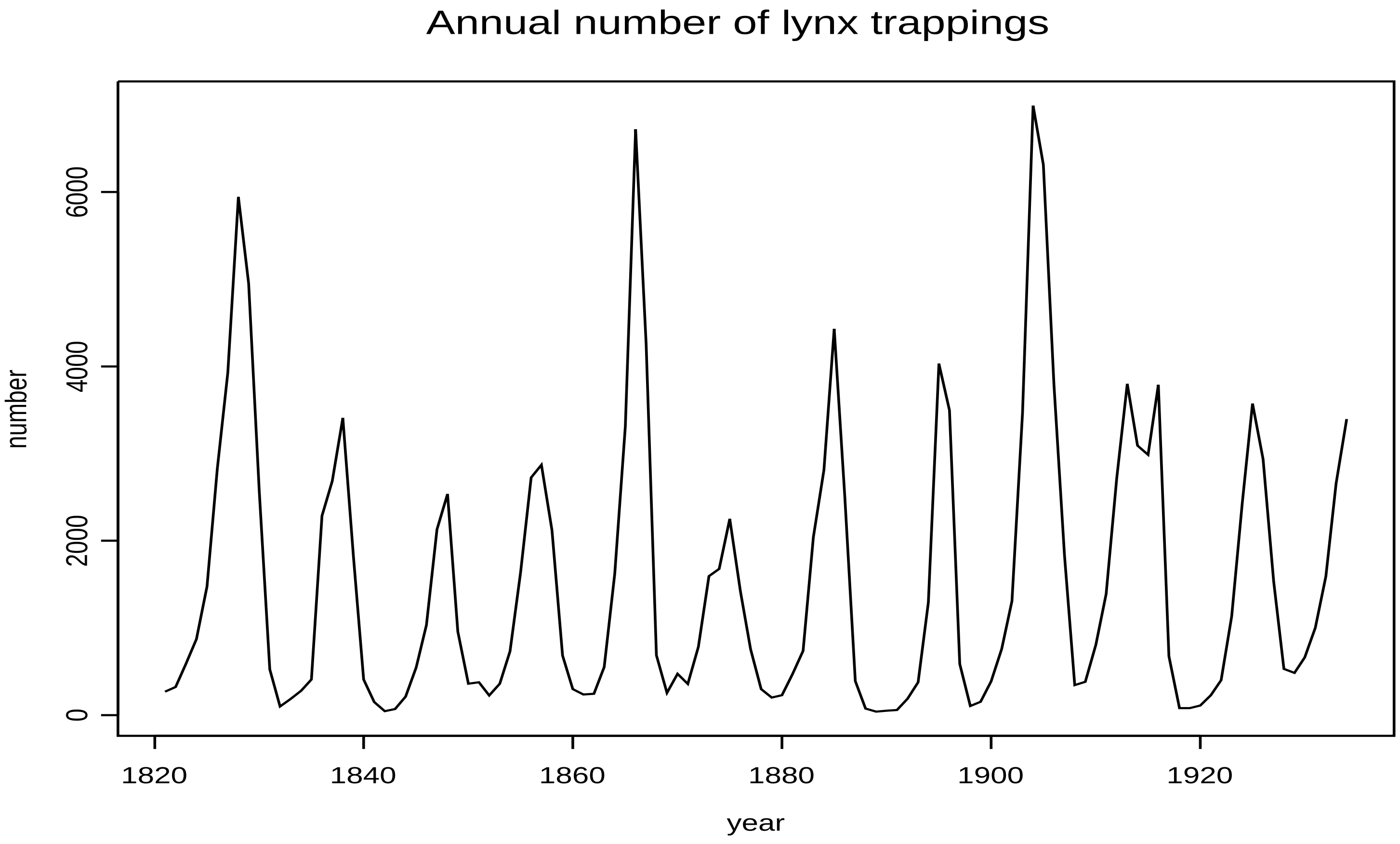
at Université Libre de Bruxelles

March 9-12, 2010

Time series data

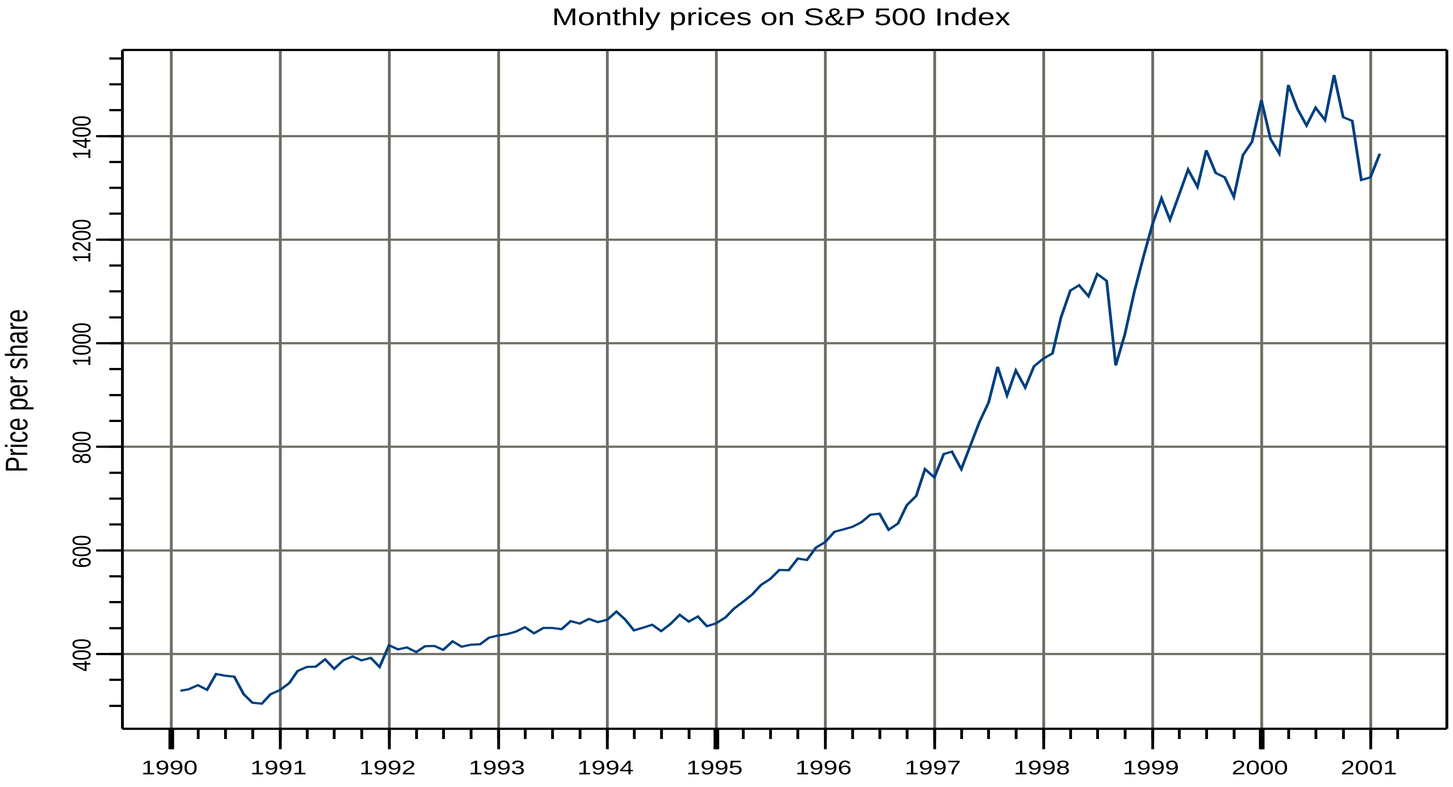
- Here we consider the empirical likelihood method for time series data.
- A time series is a sequence of observations $\boldsymbol{Y}_i \in \mathbb{R}^d$, $i = 1, \ldots, T$, where $\boldsymbol{Y}_{i+1}$ is observed one time unit after $\boldsymbol{Y}_i$. The time unit could be a fixed amount of real time, such as a day or year.

Example of the time series data



- This graph shows the number of lynx trappings in the Mackenzie River District of North-West Canada for the period 1821 to 1934. Time unit is a year.

Example of the time series data



- This graph shows the price of S&P 500 index for the period Jan 1990 to Jan 2001. Time unit is a month.

Dependency

- Models with independent $\boldsymbol{Y}_i$ seems to be inappropriate for time series data. There is generally some dependence among series values to account for.
- Empirical likelihood was originally motivated by independent identically distributed data. So we can not apply EL method to dependent data directly.
- In the following, we see the basic of time series analysis in univariate case, and after that, we consider the application of EL method for time series data.

Strictly stationary process

- A widely used assumption is that Y_i for $i = 1, \ldots, T$ are T consecutive observations from an infinite series $\ldots, Y_{-1}, Y_0, Y_1, \ldots$ having a strictly stationary distribution.

Definition (Strictly stationary)

The process $\{Y_t\}_{t\in\mathbb{Z}} \in \mathbb{R}$ is said to be strictly stationary if the joint distribution of any finite set of observations is unaffected by a time shift of k units, that is;

$$(Y_{t_1}, \ldots, Y_{t_n}) \overset{d}{=} (Y_{t_1+k}, \ldots, Y_{t_n+k})$$

for $\forall n \in \mathbb{N}$, $\forall t_1, \ldots, t_n \in \mathbb{Z}$ and $\forall k \in \mathbb{Z}$.

Weakly stationary process

- The concept of strictly stationarity is natural, but it is too strong to prescribe the joint distributions of all finite dimensions. So we consider the weaker assumtion.

Definition ((Weakly) stationary)

The process $\{Y_t\}_{t\in\mathbb{Z}} \in \mathbb{R}$ is said to be (weakly) stationary if

$$\begin{aligned}
&\text{(i) } E|Y_t|^2 < \infty \quad \forall t \in \mathbb{Z}\\
&\text{(ii) } E(Y_t) = \mu \quad \forall t \in \mathbb{Z}\\
&\text{(iii) } \Gamma_Y(t,s) = \Gamma_Y(t+k, s+k) \quad \forall t, s, k \in \mathbb{Z}
\end{aligned}$$

where $\Gamma_Y(t,s) = E\{Y_t - E(Y_t)\}\{Y_s - E(Y_s)\}$, which is called the autocovariance function.

Index of self dependency

- If $\{Y_t\}_{t\in\mathbb{Z}}$ is stationary then $\Gamma_Y(t,s) = \Gamma_Y(t-s,0)$ for $\forall t,s \in \mathbb{Z}$. It is covenient to redifine it as

$$\Gamma_Y(k) \equiv \Gamma_Y(k,0) = \mathrm{Cov}(Y_{t+k}, Y_t) \quad \forall t,k \in \mathbb{Z}$$

- In addition, we define the autocorrelation function (ACF) of $\{Y_t\}_{t\in\mathbb{Z}}$ as

$$\rho_Y(k) \equiv \frac{\Gamma_Y(k)}{\Gamma_Y(0)} = \mathrm{Corr}(Y_{t+k}, Y_t) \quad \forall t,k \in \mathbb{Z}$$

- ACF is an index which expresses the magnitude of the self dependency of the process.

Sample autocovariance function

- The sample autocovariance function of $\{Y_1, \ldots, Y_n\}$ is defined as

  $$\hat{\Gamma}(k) \equiv \frac{1}{n}\sum_{j=1}^{n-k}(Y_{j+k} - \bar{Y})(Y_j - \bar{Y}), \quad 0 \le k < n,$$

  and $\hat{\Gamma}(k) = \hat{\Gamma}_Y(-k)$, $-n < k \le 0$, where $\bar{Y}$ is the sample mean $\bar{Y} = n^{-1}\sum_{j=1}^{n} Y_j$.
- The autocovariance function is usually estimated by this sample autocovariance function.

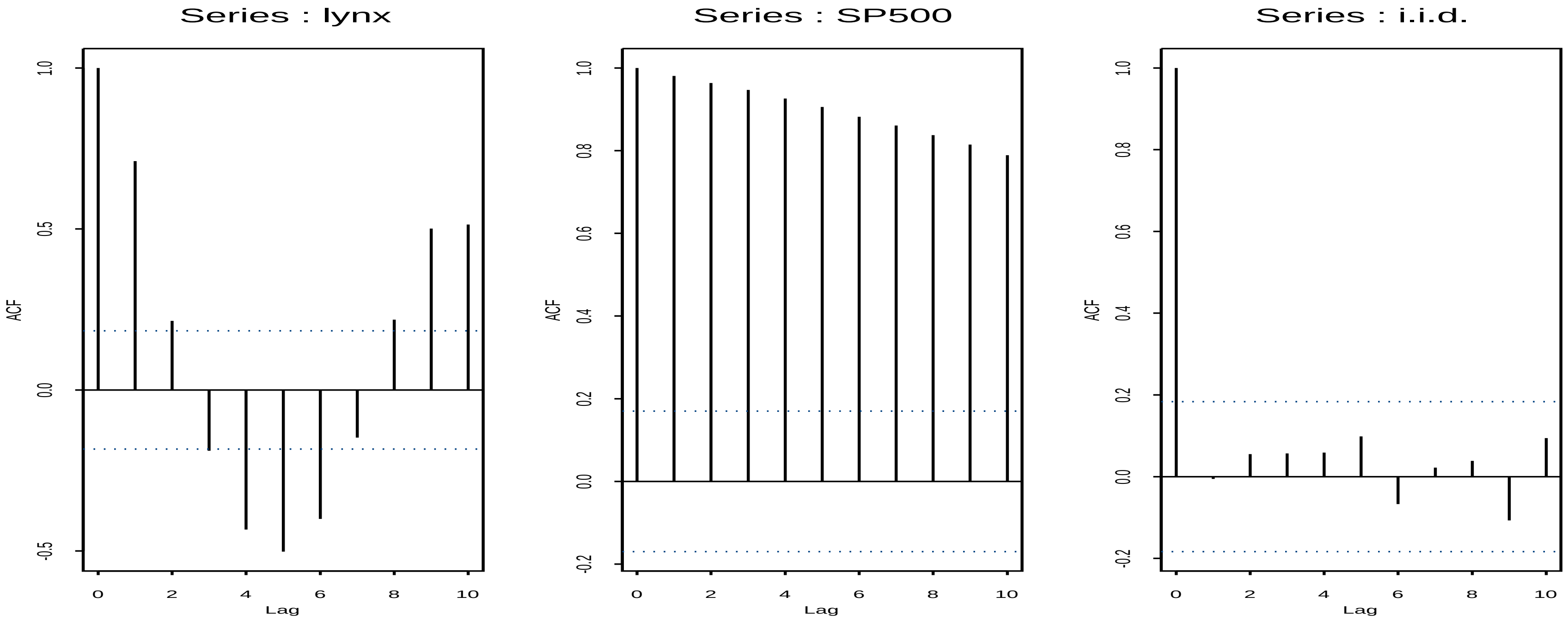
Sample autocorrelation function

- Similarly, the sample autocorrelation function of $\{Y_1, \ldots, Y_n\}$ is defined as

$$\hat{\rho}(k) \equiv \frac{\hat{\Gamma}(k)}{\hat{\Gamma}(0)}, \qquad |k| < n.$$

- The autocorrelation function is usually estimated by this sample autocorrelation function.

Sample ACFs



- From the left, sample ACFs of lynx, S&P 500 index and simulated i.i.d. data. Comparing the rightestone, we can see the left two data have strong self dependency.

Autoregressive model

- Suppose that $e_t \sim N(0, \sigma^2)$ are independent. Then let us consider the following model

  $$Y_t = \phi_0 + \phi_1 Y_{t-1} + e_t, \quad t = 1, \ldots, T \quad (*)$$

  and Y_0 is random variable independent of the e_t.
- This model is called autoregressive model and widely used in parametric modeling of time series data. As the name indicates, the data series is generated by a regression on its own past.

Autoregressive model

- Iterative calculations lead to

$$Y_t = \frac{\phi_0(1-\phi_1^t)}{1-\phi_1} + \phi_1^t Y_0 + \sum_{i=1}^{t} \phi_1^{t-i} e_i,$$

so if $|\phi_1| < 1$,

$$E(Y_t) \to \mu = \frac{\phi_0}{1-\phi_1}, \quad \mathrm{Cov}(Y_t, Y_{t+k}) \to \Gamma(k) = \frac{\phi_1^k}{1-\phi_1^2}\sigma^2.$$

Therefore in the case of $|\phi_1| < 1$, $(*)$ is an asymptoticaly stationary process.

Reducing to independence

- Assume $|\phi_1| < 1$, then we can rewrite $(*)$ as

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + e_t, \quad t = 1, \ldots, T.$$

- Now, we have three parameters $\boldsymbol{\theta} = (\mu, \phi_1, \sigma)'$. To estimate $\boldsymbol{\theta}$ with EL method, we set

$$\varepsilon_t = \varepsilon_t(\mu, \phi_1) = (Y_t - \mu) - \phi_1(Y_{t-1} - \mu)$$

  and construct the estimating function as

$$\boldsymbol{Z}_t = \boldsymbol{Z}_t(\boldsymbol{\theta}) = (\varepsilon_t, (Y_t - \mu)\varepsilon_t, \varepsilon_t^2 - \sigma^2)'.$$

Reducing to independence

- If $\boldsymbol{\theta}_0$ is the true value of the parameter then $Z_t(\boldsymbol{\theta}_0)$ are independent and satisfy

$$E[\boldsymbol{Z}_t(\boldsymbol{\theta}_0)] = \mathbf{0}, \quad t = 2, \ldots, T.$$

- The empirical likelihood approach is then based on

$$\mathcal{R}(\boldsymbol{\theta}) = \sup_{\boldsymbol{p}} \left\{ \prod_{t=1}^{n} np_t \,\middle|\, \sum_{t=1}^{n} p_t \boldsymbol{Z}_{t+1}(\boldsymbol{\theta}) = \mathbf{0}, p_t \geq 0, \sum_{t=1}^{n} p_t = 1 \right\},$$

where $n = T - 1$. If the e_i are independent $N(0, \sigma^2)$, then the limiting distribution of $-2\log\mathcal{R}(\boldsymbol{\theta}_0)$ is χ^2_3.

Reducing to independence

- In fact, the e_t do not have to be normal. They do have to be nearly independent, so that

  $$\frac{1}{T}\sum_t \boldsymbol{Z}_t'\boldsymbol{Z}_t$$

  estimates the variance matrix of

  $$\frac{1}{\sqrt{T}}\sum_t \boldsymbol{Z}_t.$$

$AR(k)$ *model*

- The autoregressive model

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + e_t.$$

  is known as the AR(1) model because it uses a regression on one past data point.

- A natural extension of AR(1) model is

$$Y_t - \mu = \sum_{j=1}^{k} \phi_j(Y_{t-j} - \mu) + e_t \quad t \in \mathbb{Z},$$

  which is called AR(k) model. Here, we just assume e_t are uncorrelated with mean 0 and constant variance σ^2.

Condition for stationary AR(k) model

- For AR(1) model, we saw that if $|\phi_1| < 1$, then the process becomes stationary.

What is the condition for AR(k) model to be stationary?

- To see this, define the polynomial

$$\phi(z) = \sum_{j=0}^{k} \phi_j z^j \quad z \in \mathbb{C}$$

where $\phi_0 = 1$.

Condition for stationary $AR(k)$ *model*

- Then, AR(k) model is rewritten as

  $$\phi(B)X_t = e_t,$$

  where $X_t = Y_t - \mu$ and B is a lag operater (i.e. $B^j X_t = X_{t-j}$).
- If we denote the roots of $\phi(z) = 0$ by $z_1, \ldots, z_k$, we can rewrite

  $$\phi(z) = \prod_{j=1}^{k} \left(1 - \frac{z}{z_j}\right)$$

Condition for stationary AR(k) model

- If $|z_j| > 1$ for all j, we can expand as

$$\left(1 - \frac{z}{z_j}\right)^{-1} = \sum_{\ell=0}^{\infty} \left(\frac{z}{z_j}\right)^{\ell},$$

and can obtain

$$\phi(z)^{-1} = \prod_{j=1}^{k} \left\{ \sum_{\ell=0}^{\infty} \left(\frac{z}{z_j}\right)^{\ell} \right\}.$$

Condition for stationary AR(k) model

- Then X_t can be expressed as a linear process of e_t;

$$X_t = \phi(B)^{-1} e_t = \prod_{j=1}^{k} \left\{ \sum_{\ell=0}^{\infty} \left(\frac{B}{z_j} \right)^{\ell} \right\} e_t = \sum_{\ell=0}^{\infty} \psi_\ell e_{t-\ell} \quad \text{(say)}.,$$

- If $|z_j| > 1$ for all j, $|\psi_\ell| \to 0$ as $\ell \to \infty$ then

$$E(X_t) = 0, \quad \mathrm{Cov}(X_t, X_{t+h}) = \sigma^2 \sum_{\ell=0}^{\infty} \psi_\ell \psi_{\ell+h} < \infty,$$

which implies stationarity of the process.

Condition for stationary AR(k) *model*

What is the condition for AR(k) model to be stationary?

$\Downarrow$ $\Downarrow$

$\Downarrow$ answer $\Downarrow$

$\Downarrow$ $\Downarrow$

- All the roots of the equation

$$\phi(z) = \sum_{j=0}^{k} \phi_j z^j = 0 \quad z \in \mathbb{C}$$

have to lie outside the unit circle in the complex plane.

Estimating function for AR(k) model

- For AR(k) model, we can construct the estimating function as

  $$Z_t = Z_t(\boldsymbol{\theta}) = (X_t, e_t X_{t-1}, \ldots, e_t X_{t-k}, e_t^2 - \sigma^2)'$$

  where $\boldsymbol{\theta} = (\mu, \phi_1, \ldots, \phi_k, \sigma)'$, $X_t = Y_t - \mu$ and $e_t = e_t(\mu, \phi_1, \ldots, \phi_k) = X_t - \sum_{j=1}^{k} \phi_j X_{t-j}$.
- The empirical likelihood approach is then based on

  $$\mathcal{R}(\boldsymbol{\theta}) = \sup_{\boldsymbol{p}} \left\{ \prod_{t=1}^{n} np_t \,\middle|\, \sum_{t=1}^{n} p_t \boldsymbol{Z}_{t+k}(\boldsymbol{\theta}) = \mathbf{0}, p_t \geq 0, \sum_{t=1}^{n} p_t = 1 \right\},$$

  where $n = T - k$.

Spectral distribution & spectral density function

- If $\{Y_t\}_{t\in\mathbb{Z}}$ is stationary then autocovariance function $\Gamma(k)$ can be expressed by the right conditinuous, nondecreasing and bouded function $F(\cdot)$ s.t.

$$\Gamma(k) = \int_{-\pi}^{\pi} e^{-i\omega k} dF(\omega), \qquad F(-\pi) = 0.$$

- $F(\cdot)$ is called spectral distribution function of $\{Y_t\}$.
- When we can write

$$\Gamma(k) = \int_{-\pi}^{\pi} e^{-i\omega k} f(\omega) d\omega,$$

then we call $f(\omega)$ the spectral density function of $\{Y_t\}$

Periodogram

- The spectral density function $f(\omega)$ is also written as

$$f(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} R(s) e^{-is\omega}$$

- The sample version of the spectral density function is defined as

$$I_T(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^{T} Y_t e^{it\omega} \right|^2 .$$

$I_T(\omega)$ is called the periodogram.

Property of the periodogram

- The periodogram $I_T(\omega)$ is an asymptotically unbiased estimator of $f(\omega)$.

$$E[I_T(\omega)] = f(\lambda) + O(T^{-1}).$$

- However, it is not a consistent estimator

$$Var(I_T(\lambda_j)) = \begin{cases} f(\lambda_j)^2 + O(T^{-1}) & j \neq 0 \\ 2f(\lambda_j)^2 + O(T^{-1}) & j = 0 \end{cases}$$

$$Cov(I_T(\lambda_j), I_T(\lambda_k)) = O(T^{-2}), \quad j \pm k \neq 0 \quad \text{mod } T$$

where $\lambda_j = \dfrac{2\pi j}{T}$, $(j \in \mathbb{Z})$ are discrete frequencies.

The case of ARMA(p,q)

- Consider the following $ARMA(p,q)$ process $\{Y_t\}$

$$\sum_{j=0}^{p} \beta_j Y_{t-j} = \sum_{j=0}^{q} \alpha_j e_{t-j} \qquad (\alpha_0 = \beta_0 = 1),$$

  where $\{e_t\} \sim IID(0, \sigma^2)$.

- The spectral density function is written as

$$f_{\boldsymbol{\theta}}(\omega) = \frac{\sigma^2}{2\pi} \frac{|\sum_{j=0}^{q} \alpha_j e^{ij\omega}|^2}{|\sum_{j=0}^{p} \beta_j e^{ij\omega}|^2}$$

  where $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_p, \alpha_1, \ldots, \alpha_q, \sigma^2) \in \Theta \subset \mathbb{R}^{p+q+1}$

Whittle likelihood

- In frequency domain, the approximate log likelihood (Whittle likelihood) can be written as

$$W(\boldsymbol{\theta}) = -\int_{-\pi}^{\pi} \left\{ \log f_{\boldsymbol{\theta}}(\omega) + \frac{I_T(\omega)}{f_{\boldsymbol{\theta}}(\omega)} \right\} d\omega$$

- Whittle's estimator $\hat{\boldsymbol{\theta}}$ maximizes $W(\boldsymbol{\theta})$ over the parameter space. Therefore

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \left\{ \log f_{\boldsymbol{\theta}}(\omega) + \frac{I_T(\omega)}{f_{\boldsymbol{\theta}}(\omega)} \right\} d\omega = \mathbf{0}$$

Whittle likelihood

- Using discrete frequency $\lambda_j = 2\pi j/T \; j = 1, \ldots T$, we construct the estimating function in frequency domain.

$$m_j(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \left\{ \log f_{\boldsymbol{\theta}}(\lambda_j) + \frac{I_T(\lambda_j)}{f_{\boldsymbol{\theta}}(\lambda_j)} \right\}$$

- Then, for true $\boldsymbol{\theta}_0$

$$E[\boldsymbol{m}_j(\boldsymbol{\theta}_0)] \to \mathbf{0}$$

So, we can construct the profile empirical likelihood ratio in frequency domain.

$$\mathcal{R}(\boldsymbol{\theta}) = \max \left\{ \prod_{t=1}^{T} T w_t | \sum_{t=1}^{T} w_t \boldsymbol{m}_j(\boldsymbol{\theta}) = \mathbf{0}, w_t \geq 0, \sum_{t=1}^{T} w_t = 1 \right\}$$