SF2943 Time Series Analysis: Lecture 7

April 28, 2022

We introduced a recursive algorithm called Durbin-Levinson algorithm for computing the best one-step linear predictor $P_n X_{n+1}$. In this lecture, we will introduce another recursive algorithm called the innovations algorithm. After that, we will focus on the sample mean and sample acvf.

1 Innovations Algorithm

Consider a time series $\{X_t\}$ (which might be non-stationary) with $EX_t = 0$ and $E|X_t^2| < \infty$ for every t. Define $\kappa(i,j) = E(X_i X_j)$,

$$\hat{X}_n = \begin{cases} 0, & \text{if } n = 1 \\ P_{n-1}X_n, & \text{if } n = 2, 3, \ldots \end{cases} \quad \text{and } \nu_n = E(X_{n+1} - P_n X_{n+1})^2,$$

$$U_n = X_n - \hat{X}_n, \text{ for any } n = 1, 2, \ldots, \tag{1.1}$$

where $P_n X_{n+1}$ is the best one-step linear predictor for X_{n+1} in terms of $\{X_1, \ldots, X_n\}$ and U_n is known as the innovation or the one-step predict error.

Using the notations $\mathbf{U}_n = (U_1, \ldots, U_n)'$ and $\mathbf{X}_n = (X_1, \ldots, X_n)'$ and recall that $\hat{X}_n = P_{n-1}X_n = \phi_{n-1,1}X_{n-1} + \cdots + \phi_{n-1,n-1}X_1$, we can rephrase (1.1) as

$$\mathbf{U}_n = A_n \mathbf{X}_n, \tag{1.2}$$

where

$$A_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{1,1} & 1 & 0 & \cdots & \vdots \\ -\phi_{2,2} & -\phi_{2,1} & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ -\phi_{n-1,n-1} & -\phi_{n-1,n-2} & -\phi_{n-1,n-3} & \cdots & 1 \end{pmatrix}.$$

Since A_n is a lower triangular matrix with non-zero values on the diagonal, A_n is invertible, with inverse C_n of the form

$$C_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 1 & 0 & \cdots & \vdots \\ \theta_{2,2} & \theta_{2,1} & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{pmatrix}.$$

The vector of one-step predictors $\hat{\mathbf{X}}_n = (X_1, P_1X_2, \ldots, P_{n-1}X_n)'$ can therefore be expressed as

$$\hat{\mathbf{X}}_n = \mathbf{X}_n - \mathbf{U}_n = C_n\mathbf{U}_n - \mathbf{U}_n = (C_n - I_n)\mathbf{U}_n = \Theta_n(\mathbf{X}_n - \hat{\mathbf{X}}_n), \tag{1.3}$$

where I_n is the $n \times n$ identity matrix and

$$\Theta_n = C_n - I_n = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 0 & 0 & \cdots & \vdots \\ \theta_{2,2} & \theta_{2,1} & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 0 \end{pmatrix}.$$

Equation (1.3) can be rewritten as

$$\hat{X}_{n+1} = \begin{cases} 0, & \text{if } n = 1 \\ \sum_{j=1}^{n} \theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right), & \text{if } n = 2, 3, \ldots \end{cases}.$$

From here we see that as long as we can compute the coefficients $\theta_{n1}, \ldots, \theta_{nn}$, then we can compute $\hat{X}_n$ with $\hat{X}_1, \ldots, \hat{X}_n$.

Theorem 1.1 (The Innovations Algorithm) *Let $\nu_0 = \kappa(1,1)$. For any $n \in \mathbb{N}$, the best linear predictor of X_{n+1} in terms of $X_1, \ldots, X_n$ is*

$$\hat{X}_{n+1} = P_nX_{n+1} = \sum_{j=1}^{n} \theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right),$$

where the coefficients $\theta_{n,1}, \ldots, \theta_{n,n}$ and ν_n can be computed recursively from $\{\theta_{k,j} : 1 \le k \le n-1, 1 \le j \le k\}$, $\{\nu_j : 1 \le j \le n-1\}$, and the equations

$$\theta_{n,n-k} = \nu_k^{-1}\left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\nu_j\right),$$

and

$$\nu_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2\nu_j.$$

2 Sample Mean and Autocorrelation Function

In order to fit our observations with a stationary time series model $\{X_t\}$, the first thing we have to do is to estimate its mean $\mu = EX_t$ and acvf $\gamma(h) = \mathrm{Cov}(X_{t+h}, X_t)$.

2.1 Estimation of μ

The estimator of the mean μ is called the sample mean and defined as

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Notice that

$$E\bar{X}_n = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} EX_i = \mu.$$

This indicates that the sample mean $\bar{X}_n$ is an unbiased estimator of the mean μ. Moreover,

$$\begin{aligned}
\mathrm{Var}(\bar{X}_n) &= E(\bar{X}_n - \mu)^2 = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2 = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right)^2 \\
&= E\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - \mu)(X_j - \mu)\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left((X_i - \mu)(X_j - \mu)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}\left(X_i, X_j\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(i-j) \\
&= \frac{1}{n^2}\sum_{h=-n}^{n}(n - |h|)\gamma(h) = \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma(h). \qquad (2.1)
\end{aligned}$$

Proposition 2.1 *If $\{X_t\}$ is a stationary time series with mean μ and acvf γ, then as $n \to \infty$,*

$$Var(\bar{X}_n) \to 0 \quad \text{if} \quad \gamma(n) \to 0$$

and

$$n\,Var(\bar{X}_n) \to \sum_{h=-\infty}^{\infty}\gamma(h) \quad \text{if} \quad \sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty.$$

Proof. From (2.1)

$$0 \le \mathrm{Var}(\bar{X}_n) = \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma(h) \le \frac{1}{n}\sum_{h=-n}^{n}\gamma(h) = \frac{1}{n}\gamma(0) + \frac{2}{n}\sum_{h=1}^{n}\gamma(h).$$

Therefore, in order to show $\mathrm{Var}(\bar{X}_n) \to 0$, it suffices to prove that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{h=1}^{n}\gamma(h) = 0.$$

If $\gamma(n) \to 0$, then for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $|\gamma(n)| \le \varepsilon$ for any $n \ge N+1$. Moreover, since

$$0 \le \frac{1}{n}\sum_{h=1}^{n}\gamma(h) = \frac{1}{n}\sum_{h=1}^{N}\gamma(h) + \frac{1}{n}\sum_{h=N+1}^{n}\gamma(h) \le \frac{1}{n}\sum_{h=1}^{N}\gamma(h) + \frac{n-(N+1)}{n}\varepsilon,$$

we find

$$0 \le \limsup_{n\to\infty}\frac{1}{n}\sum_{h=1}^{n}\gamma(h) \le \limsup_{n\to\infty}\frac{1}{n}\sum_{h=1}^{N}\gamma(h) + \limsup_{n\to\infty}\frac{n-(N+1)}{n}\varepsilon = \varepsilon.$$

Because ε is arbitrary, we know that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{h=1}^{n}\gamma(h) = 0.$$

This completes the proof for the first statement.

For the second one, by (2.1)

$$n\mathrm{Var}(\bar{X}_n) = \sum_{h=-n}^{n}\left(1-\frac{|h|}{n}\right)\gamma(h) = \sum_{h=-\infty}^{\infty}\left(1-\frac{|h|}{n}\right)I_{[-n,n]}(h)\gamma(h),$$

where $I_{[-n,n]}(h) = 0$ if $h \notin [-n,n]$ and $I_{[-n,n]}(h) = 0$ if $h \in [-n,n]$. Hence, if $\sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty$, then we can apply the dominated convergence theorem[1]

$$\lim_{n\to\infty} n\mathrm{Var}(\bar{X}_n) = \lim_{n\to\infty}\sum_{h=-\infty}^{\infty}\left(1-\frac{|h|}{n}\right)I_{[-n,n]}(h)\gamma(h) = \sum_{h=-\infty}^{\infty}\lim_{n\to\infty}\left(1-\frac{|h|}{n}\right)I_{[-n,n]}(h)\gamma(h)$$
$$= \sum_{h=-\infty}^{\infty}\gamma(h).$$

■

Remark 2.2 *Roughly speaking, the proposition says that the variance of the sample mean Var($\bar{X}_n$) decays to* 0 *as $n \to \infty$, and it decays in the rate as $1/n$. In particular,*

$$Var(\bar{X}_n) \approx \frac{1}{n}\sum_{h=-\infty}^{\infty}\gamma(h).$$

In order to determine the "performance" of the mean estimator $\bar{X}_n$, for example, to construct a confidence interval, understanding its variance is not sufficient. We have to know the distribution of $\bar{X}_n$. It turns out that $\{\bar{X}_n\}$ satisfies the central limit theorem, and therefore, we can approximate the distribution of $\bar{X}_n$ by a normal distribution.

[1] It is okay if you do not know this theorem.

Theorem 2.3 *If $\{X_t\}$ is a stationary process satisfying*

$$X_t = \mu + \sum_{j=\infty}^{\infty} \psi_j Z_{t-j}$$

where $\{Z_j\} \sim IID(0, \sigma^2)$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$, then

$$\frac{1}{\sqrt{n}}(\bar{X}_n - \mu) \xrightarrow{d} N(0, v), \text{ as } n \to \infty$$

with

$$v = \sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right).$$

Roughly speaking, the theorem says that for large n,

$$\bar{X}_n \approx N\left(\mu, \frac{1}{n} v\right)$$

with

$$v = \sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right).$$

Hence, an approximate $1 - \alpha$ confidence interval of the sample mean $\bar{X}_n$ is

$$\left(\bar{X}_n - \lambda_{\alpha/2} \sqrt{\frac{v}{n}}, \bar{X}_n + \lambda_{\alpha/2} \sqrt{\frac{v}{n}} \right),$$

where λ_α satisfies

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda_\alpha} e^{-\frac{x^2}{2}} = 1 - \alpha.$$

A special case is when $\alpha = 0.05$, then an approximate 95% confidence interval is

$$\left(\bar{X}_n - 1.96 \sqrt{\frac{v}{n}}, \bar{X}_n + 1.96 \sqrt{\frac{v}{n}} \right).$$

However, in general, v is an unknown quantity, so we have to estimate v by, for example,

$$\hat{v} = \sum_{|h| < \sqrt{n}} \left(1 - \frac{|h|}{\sqrt{n}} \right) \hat{\gamma}(h),$$

where $\hat{\gamma}(h)$ is the sample acvf.

Example 2.1 (AR(1) model) *Consider $\{X_t\}$ an $AR(1)$ process with mean μ, i.e.,*

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t$$

with $|\phi| < 1$ and $\{Z_t\} \sim WN(0, \sigma^2)$.

From our previous lectures, we know the acvf

$$\gamma(h) = \phi^{|h|}\frac{\sigma^2}{1-\phi^2}.$$

Thus,

$$v = \sum_{h=-\infty}^{\infty}\gamma(h) = \frac{\sigma^2}{1-\phi^2}\sum_{h=-\infty}^{\infty}\phi^{|h|} = \frac{\sigma^2}{1-\phi^2}\left(1+2\sum_{h=1}^{\infty}\phi^h\right) = \frac{\sigma^2}{(1-\phi)^2}$$

and approximate 95% confidence bounds for μ are given by

$$\bar{x}_n \pm 1.96\frac{\sigma}{\sqrt{n}(1-\phi)}.$$

Since ϕ and σ are unknown in practice, they must be replaced in these bounds by estimated values.

2.2 Estimation of $\gamma(\cdot)$ and $\rho(\cdot)$

Recall that the sample autocovariance function (sample acvf) is defined as

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|}-\bar{x})(x_t-\bar{x}),\ for\ -h<n<h.$$

The sample autocorrelation function (sample acf) is defined as

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)},\ for\ -h<n<h.$$

It can be shown that (see the book) the sample acvf is nonnegative definite. Therefore, the sample acvf satisfies the characterization of acvf (i.e., nonnegative definite and even).

The sample acvf $\hat{\gamma}$/acf $\hat{\rho}$ is an estimate of the acvf γ/acf ρ, but the estimation is unreliable when h is large. One can easily see this unreliability from an extreme case when $h = n-1$,

$$\hat{\gamma}(n-1) = \frac{1}{n}(x_n-\bar{x})(x_1-\bar{x}).$$

The estimation of $\hat{\gamma}(n-1)$ replies on only one quantity. A useful guide is provided by Jenkins (1976) who suggests that n should be at least about 50 and $h \leq n/4$.

Our earlier lectures discussed many different time series models and found that their acvfs behave quite differently. In particular, the acvf $\gamma(h)$ for a MA(q) process becomes 0 when $h > q$; while the acvf $\gamma(h)$ for an AR(1) process decays exponentially to 0. We also mentioned that when we plot the sample acvf $\hat{\gamma}(h)$, and if we discover that $\hat{\gamma}(h)$ becomes a small number "close" to 0 for all $h > q$ and for some $q \in \mathbb{N}$, then we could consider fitting our data with a MA(q) process. Nevertheless, how small should a number be to be considered zero? The following theorem can answer this question.

Theorem 2.4 *If $\{X_t\}$ is a stationary process satisfying*

$$X_t = \mu + \sum_{j=\infty}^{\infty} \psi_j Z_{t-j}$$

where $\{Z_j\} \sim IID(0, \sigma^2)$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and either $EZ_t^4 < \infty$ or $\sum_{j=-\infty}^{\infty} \psi_j^2 |j| < \infty$, then for each $h = 1, 2, \ldots,$

$$\frac{1}{\sqrt{n}}(\hat{\boldsymbol{\rho}}(h) - \boldsymbol{\rho}(h)) \xrightarrow{d} N(0, W), \text{ as } n \to \infty$$

with $\boldsymbol{\rho} = (\rho(1), \ldots, \rho(h))'$, $\hat{\boldsymbol{\rho}} = (\hat{\rho}(1), \ldots, \hat{\rho}(h))'$ and $W = (w_{ij})_{i,j=1}^h$ given by Bartlett's formula

$$w_{ij} = \sum_{k=1}^{\infty} (\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k))(\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)).$$

Roughly speaking, the theorem says that for large n,

$$\hat{\rho}(i) \approx N\left(\rho(i), \frac{1}{n} w_{ii}\right)$$

with

$$w_{ii} = \sum_{k=1}^{\infty} (\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k))^2.$$

So approximate 95% confidence bounds for $\rho(i)$ are given by

$$\hat{\rho}(i) \pm 1.96 \sqrt{\frac{w_{ii}}{n}}.$$

Remark 2.5 *The conditions mentioned in the theorem are satisfied by every ARMA(p, q) process driven by an iid sequence $\{Z_t\}$ with zero mean and finite variance.*

Example 2.2 (iid Noise) *If $\{X_t\} \sim IID(0, \sigma^2)$, then this means that $\phi_0 = 1$ and $\phi_j = 0$ for $j \geq 1$. Also, recall that the acvf $\rho(0) = 1$ and $\rho(h) = 0$ for $h \geq 1$. Therefore, according to Theorem 2.4,*

$$\begin{aligned} w_{ij} &= \sum_{k=1}^{\infty} (\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k))(\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)) \\ &= \sum_{k=1}^{\infty} \rho(k-i)\rho(k-j) = \begin{cases} 1, & \text{for } i = j \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

and for large n, $\hat{\rho}(1), \ldots, \hat{\rho}(h)$ are approximately independent and identically distributed normal random variables with mean 0 and variance $1/n$.

This result provides a way to test if it is reasonable to fit observed data $\{x_1, \ldots, x_n\}$ with an iid noise model. The way is to compute and plot the sample acf $\hat{\rho}(h)$. If approximately 95% of h with $\hat{\rho}(h) \in (-1.96/\sqrt{n}, 1.96/\sqrt{n})$, then this means that an iid noise model is reasonable to fit the observed data.

Example 2.3 (MA(q) process) *Consider a MA (q) process $\{X_t\}$, that is*

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

with $\{Z_t\} \sim IID(0, \sigma^2)$. From our previous lectures, $\rho(h) = 0$ for $|h| > q$, thus from Bartlett's formula, we find that for $i > q$

$$\begin{aligned} w_{ii} &= \sum_{k=1}^{\infty} (\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k))^2 = \sum_{k=1}^{\infty} \rho(k-i)^2 \\ &= 1 + 2\rho(1)^2 + \cdots + 2\rho(q)^2. \end{aligned}$$

Thus, for large n,

$$\hat{\rho}(i) \approx N\left(\rho(i), \frac{1 + 2\rho(1)^2 + \cdots + 2\rho(q)^2}{n}\right).$$