Inverse data modelling

1 Auto-regressive data model

Given I values of a time series, $x_i;\ \ i = 1, 2, ... I$, modelled as an auto-regressive process, continuously ***innovated***, we expect that we can obtain each current x_i as a weighted sum of past values plus a current innovation:

$$x_i = \sum_{j=1}^{J} a_j x_{i-j} + p_i.$$

For a purely ***resonant*** system, a sufficient number of coefficients $a_j;\ \ j = 1, 2, ... J, J < I,$ can, in principle, well describe the system. Generally, in data modelling, the continuing innovation, $p_i;\ \ i = 1, 2, ... I$, is explicitly unknown but we may either know or assume some of its statistical properties. In the simplest of cases, we may know the actual form of p_i as, for example, either an ***Dirac impulse*** or a ***Heaviside step*** function. In the case of discretely sampled data, we describe the Dirac impulse as $p_1 = 1,$ with $p_i = 0;\ \ i = 2, 3, ... I$ and the Heaviside step as $p_1 = 1;\ \ i = 1, 2, ... I$. Where the innovation is a continuous process and unknown to us except in its expected statistical properties, we describe those properties in accomplishing the inversion. In the simpler of these cases, we often assume that $p_i;\ \ i = 1, 2, ... I$ is a ***White Gaussian*** or ***purely random*** process with expected values

$$E < p_i^2 >= \sigma_p^2. \qquad E < p_i p_j >= 0; i \neq j.$$

Such a process is said to be ***uncorrelated*** with ***variance*** σ_p^2. Note that for this process $E < p_i^2 >= \sigma_p^2$ for all i; it is a ***stationary*** process which means that its statistical properties do not evolve with time or index. In modelling with this innovation, we seek that innovation that fully describes the data having the ***minimum*** variance. If we have some knowledge that the innovation is not purely random but rather has values that correlate with each other, we introduce the correlation through a ***correlation matrix***, sometimes called the ***variance-covariance matrix***. For any n, m, this matrix is formed with its $m,\ n$ element as $E < p_m p_n >$. For real-valued innovation, it is a symmetric matrix, for complex-valued innovation, it is conjugate-symmetric. The correlation matrix, $\overline{\mathbf{V}}$, that describes an uncorrelated innovation reduces to the ***identity matrix***, $\overline{\mathbf{I}}$ as $E < p_i p_j >\equiv 0$, for $i \neq j$. For a White Gaussian innovation we can form an estimate of $E < p_i^2 >$ as

$$\hat{E} < p_i^2 > = \ \frac{1}{I} \vec{p}^{\,T} \cdot \overline{\mathbf{I}} \cdot \vec{p}$$

with the vector $\vec{p}^{\,T} = [\, p_1\ \ p_2\ \ p_3 ...\ \ p_I]$ being the ordered I values of the innovation.

Writing the autoregression equation, above, in linear algebraic form while noting that the first predictable value is x_{J+1}:

$$\begin{bmatrix} x_{J+1} \\ x_{J+2} \\ x_{J+3} \\ . \\ . \\ . \\ x_I \end{bmatrix} = \begin{bmatrix} x_J & x_{J-1} & x_{J-2} & . & . & x_1 \\ x_{J+1} & x_J & x_{J-1} & . & . & x_2 \\ x_{J+2} & x_{J+1} & x_J & . & . & x_3 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ x_{I-1} & x_{I-2} & x_{I-3} & . & . & x_{I-J} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ . \\ a_J \end{bmatrix} + \begin{bmatrix} p_{J+1} \\ p_{J+2} \\ p_{J+3} \\ . \\ . \\ p_I \end{bmatrix}$$

Reforming this vector-matrix equation symbolically:

$$\vec{x} = \overline{\mathbf{X}} \cdot \vec{a} + \vec{p}.$$

We seek the vector of autoregression coefficients, $\vec{a}$, that minimizes the expected value of the innovation energy, $\vec{p}^{\,T}\,\vec{p}$, essentially the sum of all squared values of p_i; $i = J+1,\ J = 2,\ \ldots I$.

Form, now,

$$\vec{p} = \vec{x} - \overline{\mathbf{X}} \cdot \vec{a},$$

so that

$$\vec{p}^{\,T}\,\vec{p} = (\vec{x}^{\,T} - \vec{a}^{\,T} \cdot \overline{\mathbf{X}}^{\,T}) \cdot (\vec{x} - \overline{\mathbf{X}} \cdot \vec{a})$$

$$= \vec{x}^{\,T} \cdot \vec{x} - \vec{x}^{\,T} \cdot \overline{\mathbf{X}} \cdot \vec{a} - \vec{a}^{\,T} \cdot \overline{\mathbf{X}}^{\,T} \cdot \vec{x} + \vec{a}^{\,T} \cdot \overline{\mathbf{X}}^{\,T} \cdot \overline{\mathbf{X}} \cdot \vec{a}.$$

Minimizing $\vec{p}^{\,T}\,\vec{p}$ with respect to $\vec{a}^{\,T}$, noting that $\partial \vec{a} / \partial \vec{a}^{\,T} = \mathbf{0}$, we obtain

$$\frac{\partial (\vec{p}^{\,T}\,\vec{p})}{\partial \vec{a}^{\,T}} = -\overline{\mathbf{X}}^{\,T} \cdot \vec{x} + \overline{\mathbf{X}}^{\,T} \cdot \overline{\mathbf{X}} \cdot \vec{a} = \mathbf{0}.$$

Solve for

$$\vec{a} = (\overline{\mathbf{X}}^{\,T} \cdot \overline{\mathbf{X}})^{-1}\, \overline{\mathbf{X}}^{\,T} \cdot \vec{x}.$$

The innovation is often called the ***prediction error*** and the J-coefficient series a_j, the ***prediction error filter*** or ***PEF***. Knowing the a_j we can ***deconvolve*** our data series x_i to find the innovation. We may regard this innovation as that unknown series that excites the autoregression operator that describes the ***linear system*** of observation.

We can extend this development to deal with self-correlated innovations to find, then,

$$\vec{a} = (\overline{\mathbf{X}}^{\,T} \cdot \overline{\mathbf{V}} \cdot \overline{\mathbf{X}})^{-1}\, \overline{\mathbf{X}}^{\,T} \cdot \overline{\mathbf{V}} \cdot \vec{x}.$$

We may factor $\overline{\mathbf{V}} = \overline{\mathbf{U}}^{\mathrm{T}} \cdot \overline{\mathbf{U}}$ where $\overline{\mathbf{U}}$ is ***upper diagonal*** matrix and, hence, $\overline{\mathbf{U}}^{\mathrm{T}}$, ***lower diagonal***. We, then minimize the $\vec{p}^{\,T}\overline{\mathbf{U}}^{\mathrm{T}} \cdot \overline{\mathbf{U}} \cdot \vec{p}$ with respect to $\vec{a}^{\,T}$ as before.

You may note that we can write the equation in solution for $\vec{a}$ immediately preceding

$$\vec{a} = (\overline{\mathbf{Y}}^{T} \cdot \overline{\mathbf{Y}})^{-1} \overline{\mathbf{Y}}^{T} \cdot \vec{y}$$

where, now, $\overline{\mathbf{Y}} = \overline{\mathbf{U}} \cdot \overline{\mathbf{X}}$ and $\vec{y} = \overline{\mathbf{U}} \cdot \vec{x}$. That is, we ***pre-filter*** our data with coefficients determined by those of the matrix $\overline{\mathbf{U}}$. A rougher, perhaps *ad hoc*, approach to accomplishing this is to filter the raw data series $x_i,\ i = 1, 2, \ldots I$ with a filter that would produce the equivalent of a White Gaussian process from that of the innovation process that we actually have or seek. For example, if our data model were to be innovated by a ***Brownian process***, we could simply differentiate (by either running or central ***finite differences***) our data series.

Note, the development above holds as well for complex-valued data series. One need only recall that for complex vectors and matrices, one uses ***conjugate transpose*** wherever ***transpose*** appears in the linear algebra above.

2 Auto-regressive spectral analysis

In classical AR modelling, we separate our time series into an excitation (the innovation) and the system model (described by our AR coefficients). We have modelled our innovation as minimum power white Gaussian noise. Then all the spectral character of the datum time series must be described by the system function because the innovation is already colourless/flat/white. We can determine the power density spectrum of the time series knowing our sample innovation energy, $\vec{\hat{p}}^{\,T}\,\vec{\hat{p}}$ which we can determine as

$$\vec{\hat{p}} = \vec{x} - \overline{\mathbf{X}} \cdot \vec{\hat{a}}$$

where $\vec{\hat{a}}$ is our solution vector. We determine the power density spectrum as:

$$S_x(f) = \frac{\frac{1}{I-J} \cdot \vec{\hat{p}}^{\,T}\,\vec{\hat{p}}}{\mathscr{F}^2\{1 - \vec{\hat{a}}\}}.$$

The symbol $\mathscr{F}$ represents the Fourier transform, the discrete Fourier transform in the case of a digital sequence. The argument of this Fourier transform is a K-length sequence

$$\{1, -\hat{a}_1, -\hat{a}_2, -\hat{a}_3, \ldots - \hat{a}_J, 0, 0\ 0, \ldots\},$$

where $K \geq J + 1$. While all the available spectral information is obtained with $K = J + 1$, it is usually seen to show a more "attractive" spectrum if $K \gg J + 1$. One might reasonably choose a K convenient for the scaling of the discrete Fourier transform one is using and for the apparent resolution of the resulting power density spectrum, $S_x(f)$ desired.

3 Auto-regressive-integrated modelling

It is not uncommon that a time series is better considered as being innovated by a random walk or brown series. It is common to model, for example, stock-market indices in this manner. The random walk models the apparent accumulating evolutionary increase or decrease in numerical values of market indices. In market modelling, we often seek to find the $\vec{\hat{a}}$-predictor that might be inherent in the market runs. The brown character of the presumed innovation is removed by creating a sequence of back-differences:

$$x'_i = x_i - x_{i-1}.$$

We, then, model as above (classical AR-inversion) for a white Gaussian innovation using the difference sequence $x'_i, \;\; i = 2, 3, \ldots I$.

If we are interested in the power density spectrum of the original sequence, $x_i \;\; i = 1, 2, 3, \ldots I$, we would simply compute the $S_x(f)$ as above and rescale it frequency-by-frequency by multiplying each frequency obtained by $1/f^2$. Perhaps it is clear that in so doing, we face a singularity at $f = 0$. This arises as we have implicitly violated a strict condition of power spectral analysis: the sequence must be a sample of a "stationary process". By that we mean that the process from which our sample sequence is derived must have finite variance. A random walk explicity fails this condition as its variance increases with the length of the process. As a work around, we would normally avoid rescaling for frequencies near 0. That is we band-limit our spectrum for $f > f_{min}$.

In principle, what we did in obtaining back differences was to obtain a pre-filtered version of our original brown-innovated sequence. This back differencing removes the accumulation or integration inherent in the brown innovation. This differencing filter handles brown innovation. What of flicker-noise innovation for which the spectral weighting from the innovation obtains $S_p f \sim f$ rather than $S_p f \sim f^2$ as in the case of brown innovation? We would need to describe a pre-filter process that obtained a power density scaling by $1/f$; the back differencing scales by $1/f^2$. We can design such filters but they tend to be quite long in numbers of coefficients if we are to obtain a good representation of the spectrum of power densities.

4 Complex-valued data

All the linear algebraic development above works directly for complex-valued data series. Geophysical, physical and electrical time-series are very often most easily described in terms of complex-valued time series.

The only accommodation in the linear algebraic description of the AR-inverse theory developed so far is to note that wherever one has used the transpose of a vector or matrix, one must use the conjugate transpose. Often the conjugate transpose is noted with a † (dagger)

rather than the capitalized *T*. So, for example in the case of an expectation of uncorrelated, minimum variance innovation, $\vec{p}^{\dagger}\vec{p}$ minimum,

$$\vec{a} = (\overline{\mathbf{Y}}^{\dagger} \cdot \overline{\mathbf{Y}})^{-1} \overline{\mathbf{Y}}^{\dagger} \cdot \vec{y}.$$