Discrete time processes

Predictions are difficult. Especially about the future

Mark Twain.

Florian Herzog

2013
Modeling observed data

When we model observed (realized) data, we encounter usually the following situations:

- Time-series data of a single time series \( (D_t) \), e.g. economic data.
- Online observations also single time series \( (D_t) \)
- Ensemble data from measurements (experiments) or seasonal data \( (D^j_t) \)

Modeling ensemble data is often easier than single time series, because we have the same stochastic information of a given point in time \( t \) with multiple realizations.

By \( D_t \) \( t = 1, \ldots, T \) we denote the data observation at time in the case that we have only one single time series observation. By \( D^j_t \) \( t = 1, \ldots, T \) \( j = 1, \ldots, l \) we denote the time series observation at time \( t \) and measurement \( j \).

Data \( D_t \) is recorded in discrete intervals (sampled data) and assumed to be generated by an SDE.
Modeling real-world data

Stochastic differential equations are can be viewed as stochastic process from two point of view:

- Stochastic process with a probability distribution $p(t|\theta)$ across-time, where $\theta$ denotes the parameters
- A stochastic process, where the changes in the resulting time series is the stochastic process, i.e. $p(D_t - D_{t-1}|\theta)$ or $p\left(\frac{D_t - D_{t-1}}{D_{t-1}}|\theta\right)$

The first interpretation is help full to describe ensemble data and the second to analyze single time series.

In order to deal with discrete data, all SDEs need to be discretized.
Discrete-time white noise

When we assume a discrete white noise process $\varepsilon_t$ which identically and independently distributed (i.i.d) with zero expectation and $\sigma^2$ variance. Then following statistical properties apply:

$$E[\varepsilon_t] = 0 \quad \text{(1)}$$
$$E[\varepsilon_t^2] = \sigma^2 \quad \text{(2)}$$
$$E[\varepsilon_t \varepsilon_\tau] = 0 \text{ for } t \neq \tau \quad \text{(3)}$$

The last property is derived from the assumption of independence. A discretized Brownian Motion is a discrete-time white noise process with $\varepsilon_t \sim \mathcal{N}(0, \Delta t)$ where $\Delta t$ is the time discretization.

In discrete-time models, a white noise process can be normally distributed (Gaussian white noise) but can be distributed by any other distribution as long as the i.i.d. assumption is valid.
Stationarity: A process is covariance-stationary if and only if (iff):

\[ E[D_t] = \mu \ \forall t \]  \hspace{1cm} (4)

\[ E[(D_t - \mu)(D_{t-j} - \mu)] = \nu_j \ \forall t \ \forall j. \]  \hspace{1cm} (5)

This form of stationarity is often also called weak stationarity. For example the process: 
\[ D_t = \mu + \varepsilon_t \] where \( \varepsilon_t \) is white noise process with zero mean and \( \sigma^2 \) variance. In this case \( D_t \) is stationary since:

\[ E[D_t] = \mu \]

\[ E[(D_t - \mu)(D_{t-j} - \mu)] = \begin{cases} \sigma_t & j = 0 \\ 0 & j \neq 0 \end{cases} \]
Statistical Theory

Stationarity: A process is strictly stationary iff for any value of $j_1, j_2, j_3, \ldots$ the joint distribution of $D_t, D_{t+j_1}, D_{t+j_2}, D_{t+j_3}, \ldots$ depend only on the intervals separating the different values of $j$ and not the time $t$.

Ergodicity: A process is called ergodic for the mean iff:

$$\lim_{t \to \infty} E[D_t] = \mu .$$

A process is said to be ergodic for the second moment iff:

$$\lim_{T \to \infty} \frac{1}{T-j} E[(D_t - \mu)(D_{t-j} - \mu)] \to^p \nu_j$$

where $\to^p$ denotes convergence in probability. From an ergodic process, we can estimate all necessary data without possing ensemble data. If an process is not ergodic, we have only one observation, we can not identify its parameters.
Statistical Theory

It can be shown that if the autocovariance of a covariance stationary process satisfies the following:

$$\sum_{j=0}^{\infty} |\nu_j| < \infty$$

then the process is ergodic for the mean.
First Order Moving Average Process

When $\varepsilon_k$ be a Gaussian white noise processes with mean zeros and variance $\sigma^2$ and consider the process:

$$ Y_k = \mu + \varepsilon_k + \theta \varepsilon_{k-1}, $$  \hspace{1cm} (9)

where $\mu$ and $\theta$ are any real valued constant. The process is called first order moving average process and usually denotes as MA(1). The expectation of $Y_k$ is given as:

$$ E[Y_k] = E[\mu + \varepsilon_k + \theta \varepsilon_{k-1}] $$

$$ = \mu + E[\varepsilon_k] + \theta E[\varepsilon_{k-1}] = \mu $$

The constant $\mu$ is thus the mean of the processes.
The variance of the MA(1) is calculated as

\[ \text{Var}[Y_k] = E[(Y_k - \mu)^2] = E[(\epsilon_k + \theta \epsilon_{k-1})^2] \]
\[ = E[\epsilon_k^2 + \theta^2 \epsilon_{k-1}^2 + 2\theta \epsilon_k \epsilon_{k-1}] \]
\[ = \sigma^2 + 0 + \theta^2 \sigma^2 = (1 + \theta^2)\sigma^2 \]

The first autocovariance is computed as

\[ E[(Y_k - \mu)(Y_{k-1} - \mu)] = E[(\epsilon_k + \theta \epsilon_{k-1})(\epsilon_{k-1} + \theta \epsilon_{k-2})] \]
\[ = E[\epsilon_k \epsilon_{k-1} + \theta \epsilon_{k-1}^2 + \theta \epsilon_{k-1} \epsilon_k \theta \epsilon_{k-1} \epsilon_{k-2}] \]
\[ = 0 + \theta \sigma^2 + 0 + 0 = \theta \sigma^2 \]
First Order Moving Average Process

The higher order autocovariances are all zero:

\[ E[(Y_k - \mu)(Y_{k-j} - \mu)] = E[(\varepsilon_k + \theta\varepsilon_{k-1})(\varepsilon_{k-j} + \theta\varepsilon_{k-j-1})] = 0. \]

Since the mean and autocovariance are not function of times, an MA(1) process is covariance stationary regardless of the value of \( \theta \), since

\[ \sum_{j=0}^{\infty} |\nu_j| = (1 + \theta^2)\sigma^2 + |\theta\sigma^2|. \] (10)

When \( \varepsilon_k \) is a Gaussian white noise process, the MA(1) is ergodic for all moments.
First Order Moving Average Process

Autocorrelation is defined as $\rho_1 = \frac{Cov[Y_k,Y_{k-1}]}{\sqrt{Var[Y_k]Var[Y_{k-1}]}}$ and $Var[Y_k] = Var[Y_{k+1}] = (1 + \theta^2)\sigma^2$. Therefore we get

$$\rho_1 = \frac{\theta \sigma^2}{(1 + \theta^2)\sigma^2} = \frac{\theta}{1 + \theta^2}. \quad (11)$$

All higher order autocorrelations are zero. Positive values of $\theta$ induce positive autocorrelation, negative values induces negative autocorrelation. The autocorrelation of an MA(1) process can be between -0.5 and 0.5. For any value of $\rho_1$, there exists two values of $\theta$, see

$$\rho_1 = \frac{(1/\theta)}{1 + (1/\theta)^2} = \frac{\theta^2(1/\theta)}{\theta^2(1 + (1/\theta)^2)} = \frac{\theta}{1 + \theta^2}. \quad (12)$$
A q-th order moving average process, denoted as MA(q) is given by

\[ Y_k = \mu + \varepsilon_k + \theta_1 \varepsilon_{k-1} + \theta_2 \varepsilon_{k-2} + \ldots \ldots + \theta_q \varepsilon_{k-q} , \]  

(13)

where \( \mu \) and \( \theta_1, \ldots, \theta_q \) are real valued constants. The expectation of \( Y_k \) is given as:

\[ E[Y_k] = \mu + E[\varepsilon_k] + \theta_1 E[\varepsilon_{k-1}] + \ldots \ldots + \theta_q E[\varepsilon_{k-q}] = \mu \]

The constant \( \mu \) is thus the mean of the processes. The variance is computed as The variance of the MA(1) is calculated as

\[ \text{Var}[Y_k] = E[(Y_k - \mu)^2] = E[(\varepsilon_k + \theta_1 \varepsilon_{k-1} + \theta_2 \varepsilon_{k-2} + \ldots \ldots + \theta_q \varepsilon_{k-q})^2] . \]

Since all \( \varepsilon \)'s are uncorrelated, all cross terms \( \varepsilon_j \varepsilon_k \) are zero.
Q-th Order Moving Average Process

The variance is therefore

$$\text{Var}[Y_k] = \sigma^2 + \theta_1^2 \sigma^2 + \ldots + \theta_q^2 \sigma^2$$

$$= \sigma^2 (1 + \sum_{j=1}^{q} \theta_j^2).$$

The autocovariance is for the i-th lag is computed as

$$\nu_i = \mathbb{E}[(\varepsilon_k + \sum_{j=1}^{q} \theta_j \varepsilon_{k-j})(\varepsilon_{k-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{k-j-i})]$$

$$= \mathbb{E}[\theta_i \varepsilon_{k-i}^2 + \theta_{i+1} \theta_1 \varepsilon_{k-i-1}^2 + \theta_{i+2} \theta_2 \varepsilon_{k-i-2}^2 + \ldots + \theta_q \theta_{q-i} \varepsilon_{k-q}^2]$$
Q-th Order Moving Average Process

The $\epsilon$'s at different dates have been dropped since their product is zero. For $i > q$ there are no $\epsilon$'s with common dates in the definition of $\nu_i$. We get for the autocovariance thus

$$\nu_i = \begin{cases} (\theta_i + \theta_{i+1}\theta_1 + \theta_{i+2}\theta_2 + \ldots + \theta_q\theta_{q-j})\sigma_i^2 & i = 1, 2, \ldots, q \\ 0 & i > q \end{cases}$$

The autocorrelation is given as

$$\nu_i = \begin{cases} \frac{\theta_i + \theta_{i+1}\theta_1 + \theta_{i+2}\theta_2 + \ldots + \theta_q\theta_{q-j}}{(1 + \sum_{j=1}^q \theta_j^2)} & i = 1, 2, \ldots, q \\ 0 & i > q \end{cases}$$
**Autoregressive first order Process**

We analyze an so-called first order autoregressive (AR(1)):

\[ Y_{k+1} = c + \Phi Y_k + \varepsilon_k \]

Long-term mean: We take the mean on both sides of the equation

\[
\begin{align*}
E[Y_{k+1}] &= E[c + \Phi Y_k + \varepsilon_k] \\
E[Y_{k+1}] &= c + \Phi E[Y_k] \\
\mu &= c + \Phi \mu \\
\mu &= \frac{c}{1 - \Phi}
\end{align*}
\]
Autoregressive first order Process

\[ \mu = \frac{c}{1 - \Phi} \]

From this formula, we can see that \( \Phi \) must be smaller than 1, otherwise the mean would be negative when \( c \geq 0 \). Also we have assumed that \( \text{E}[Y_k] = \text{E}[Y_{k+1}] = \mu \) and exists. Second moment:

\[
Y_{k+1} = \mu (1 - \Phi) + \Phi Y_k + \varepsilon_k \\
Y_{k+1} - \mu = \Phi (Y_k - \mu) + \varepsilon_k
\]
Autoregressive first order Process

\[
\begin{align*}
E[(Y_{k+1} - \mu)^2] &= E[(\Phi(Y_k - \mu) + \varepsilon_k)^2] \\
&= \Phi^2 E[(Y_k - \mu)] + 2\Phi E[(Y_k - \mu)\varepsilon_k] + E[\varepsilon_k^2] \\
E[(Y_{k+1} - \mu)^2] &= \Phi^2 E[(Y_k - \mu)] + \sigma^2
\end{align*}
\]

where \( E[(Y_k - \mu)\varepsilon_k] = 0 \). Assuming covariance stationarity, i.e. \( E[(Y_{k+1} - \mu)^2] = \nu \) and \( E[(Y_k - \mu)^2] = \nu \). We get:

\[
\nu = \Phi^2 \nu + \sigma^2
\]

\[
\nu = \frac{\sigma^2}{1 - \Phi^2}
\]
Autoregressive first order Process

Autocorrelation:

\[
E[(Y_{k+1} - \mu)(Y_k - \mu)] = E[(\Phi(Y_k - \mu) + \varepsilon_k)(Y_k - \mu)] \\
E[(Y_{k+1} - \mu)(Y_k - \mu)] = \Phi E[((Y_k - \mu)^2) + E[(\varepsilon_k)(Y_k - \mu)].
\]

Again \(E[(\varepsilon_k)(Y_k - \mu)] = 0\) and \(E[((Y_k - \mu)^2) = \nu\) we get:

\[
E[(Y_{k+1} - \mu)(Y_k - \mu)] = \Phi \nu = \Phi \frac{\sigma^2}{1 - \Phi^2}.
\]

Autocorrelation is defined as \(Cov[Y_{k+1}Y_k]/\sqrt{Var[Y_{k+1}]Var[Y_k]}\) and \(Var[Y_k] = Var[Y_{k+1}] = \nu\). Therefore we get

\[
acf(Y_{k+1}Y_k) = \phi
\]  \hspace{1cm} (14)
Autoregressive first order Process

AR(1) process with $Y_{k+1} = \Phi Y_k + c + \varepsilon_k$ has the following properties:

**Mean:** $\mu = \frac{c}{1 - \Phi}$

**Variance:** $\nu = \frac{\sigma^2}{1 - \Phi^2}$

**Autocorr.:** $acf(Y_{k+j}Y_k) = \phi^j$

A covariance stationary AR(1) process is only well defined when $|\Phi| \leq 1$, otherwise the variance is infinite and mean makes no sense. We can check if a process is a AR(1) process by calculating the mean and variance over-time and the autocorrelation. From these three calculations the parameters can be estimated. It is important to notice that the autocorrelation is exactly the coefficient of the autoregression.
Example: GBM and mean reversion

The two examples have a discrete-time structure as an first-order autoregressive process with:

\[(I) \quad \Phi = 1 + \mu \Delta t\]
\[(II) \quad \Phi = 1 - b \Delta t\]

When we assume that \(\mu > 0\), the first model is instable and we can not use the techniques from auto-regressions to estimate the parameters. We have to deal differently in this case, by using \(p = \ln(x)\), since then we get an SDE with constant coefficients.

When we assume that \(b > 0\), the process is a true mean reverting process that can be estimated by AR(1) methods. In this case, the discrete-time model is also stable.
A first order difference equation given by

\[ y_{k+1} = \Phi y_k + w_k \]

is of the same type as an AR(1) equation. The can be solved by recursive substitution since
\[ y_1 = \Phi y_0 + w_0, \quad y_2 = \Phi y_1 + w_1, \quad y_3 = \Phi y_2 + w_2 \ldots \]

When we plug in the results of \( y_1 \) into the equation for \( y_2 \), we get
\[ y_2 = \Phi^2 y_0 + \Phi w_0 + w_1 \]

which in turn we plug in the equation for \( y_3 \) and get
\[ y_3 = \Phi^3 y_0 + \Phi^2 w_0 + \Phi w_1 + w_2 \]

Expanding this further on, we get the general solution

\[ y_{k+1} = \Phi^{k+1} y_0 + \sum_{j=0}^{k} \Phi^{k-j} w_j \]
A p-th order difference equation give by

\[ y_{k+1} = \Phi_1 y_k + \Phi_2 y_{k-1} + \ldots + \Phi_p y_{k-p+1} + w_k \]

We write this in a vectorized form with:

\[
\begin{bmatrix}
    y_k \\
    y_{k-1} \\
    y_{k-2} \\
    \vdots \\
    y_{k-p-1}
\end{bmatrix}
= \begin{bmatrix}
    \Phi_1 & \Phi_2 & \Phi_3 & \cdots & \Phi_{p-1} & \Phi_p \\
    1 & 0 & 0 & \cdots & 0 & 0 \\
    0 & 1 & 0 & \cdots & 0 & 0 \\
    \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
    0 & 0 & 0 & \cdots & 1 & 0
\end{bmatrix}
\begin{bmatrix}
    w_k \\
    0 \\
    0 \\
    \vdots \\
    0
\end{bmatrix}
\]

and get

\[ \xi_{k+1} = F \xi_k + v_k. \]
The solution is given as:

\[ \xi_k = F^k \xi_0 + F^{k-1} v_0 + F^{k-2} v_1 + F^{k-3} v_2 + \ldots + F^{v_{k-1}} + v_k \]

The p-th order difference equation is stable (or stationary) when the eigenvalues of \( F \) are smaller than one. The eigenvalue polynom is given as:

\[ \lambda^p - \Phi_1 \lambda^{p-1} - \Phi_2 \lambda^{p-2} - \ldots - \Phi_{p-1} \lambda - \Phi_p = 0 \]
A highly useful operator for time series models is the lag operator. Suppose that we start with a sequence of data $x_k$ where $k = -\infty \ldots \infty$ and generate a new sequence $y_k$ where the value of $y$ for date $k$ is equal to the value $x$ took at time $k - 1$

$$y_k = x_{k-1}.$$ 

This is described as applying the Lag operator on $x_k$ and is represented by symbol $L$:

$$Lx_k = x_{k-1}.$$ 

Consider applying the lag operator twice to a series:

$$L(Lx_k) = L(x_{k-1}) = x_{k-2},$$

which is denotes as $L^2$ or $L^2 x_k = x_{k-2}$. 
Lag operators

In general for any integer $l$

$$L^l x_k = x_{k-l},$$

The lag operator is linear and can be treated as the multiplication operator and thus

$$L(\beta x_k) = \beta Lx_k = \beta x_{k-1}$$

$$L(x_k + w_k) = Lx_k + Lw_k$$

$$y_k = (a + bL)Lx_k = (aL + bL^2)x_k = ax_{k-1} + bx_{k-2}$$

An expression such as $(aL + bL^2)$ is referred as a polynomial lag operators.
Notice if $x_k = c$ then $Lx_k = x_{k-1} = c$
First order difference equation and lag operators

A first order difference equation given by

\[ y_{k+1} = \Phi y_k + w_k \]

can be written as

\[ y_{k+1} = \Phi Ly_{k+1} + w_k \]

\[ y_{k+1} - \Phi Ly_{k+1} = w_k y_{k+1} (1 - L) = w_k \]

when we multiply both sides with \((1 + \Phi L + \Phi^2 L^2 + \Phi^3 L^3 + \ldots + \Phi^k L^k)\) we get

\[ y_{k+1} (1 + \Phi L + \Phi^2 L^2 + \Phi^3 L^3 + \ldots + \Phi^k L^k) (1 - L) = (1 + \Phi L + \Phi^2 L^2 + \Phi^3 L^3 + \ldots + \Phi^k L^k) w_k \]
First order difference equation and lag operators

\[ y_{k+1}(1 - \Phi^k L^k) = (1 + \Phi L + \Phi^2 L^2 + \Phi^3 L^3 + \ldots + \Phi^k L^k)w_k \]

\[ y_{k+1}(1) = \Phi^k L^k y_{k+1} + (1 + \Phi L + \Phi^2 L^2 + \Phi^3 L^3 + \ldots + \Phi^k L^k)w_k \]

\[ y_{k+1} = \Phi^k y_0 + w_k + \Phi w_{k-1} + \Phi^2 w_{k-2} + \Phi^3 w_{k-3} + \ldots + \Phi^k w_0 \]

which results exactly in the same solution as by the usage of the recursive substitution.
P-th order difference equation and lag operators

The p-th order difference equation given by

\[ y_{k+1} = \Phi_1 y_k + \Phi_2 y_{k-1} + \ldots + \Phi_p y_{t-p+1} + w_k , \]

can be written using the lag operator as:

\[ y_{k+1}(1 - \Phi_1 L - \Phi_2 L^2 - \Phi_3 L^3 - \ldots - \Phi_p L^p) = w_k , \]

which results in the same polynomial as in the case of the eigenvalues of the matrix \( F \).
P-th order difference equation and lag operators

Since

\[(1 - \Phi_1 z - \Phi_2 z^2 - \Phi_3 z^3 - \ldots - \Phi_p z^p) = (1 - \lambda_1 z)(1 - \lambda_2 z)\ldots(1 - \lambda_p z),\]

and by defining \(z^{-1} = \lambda\) and multiply both sides with \(z^{-p}\) and we get

\[(\lambda^p - \Phi_1 \lambda^{p-1} - \Phi_2 \lambda^{p-2} - \ldots - \Phi_p) = (\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_p z),\]

which is nothing but the root search of the left side polynomial:

\[\lambda^p - \Phi_1 \lambda^{p-1} - \Phi_2 \lambda^{p-2} - \ldots - \Phi_{p-1} \lambda - \Phi_p = 0\]
Second order Autoregression

A second order autoregressive process (denotes as AR(2)) has the form:

$$Y_{k+1} = c + \Phi_1 Y_k + \Phi_2 Y_{k-1} + \varepsilon_k$$

written in lag notation:

$$Y_{k+1}(1 - \Phi_1 L - \Phi_2 L^2) = c + \varepsilon_k$$

The AR(2) is stable if the roots of $$(1 - \Phi_1 z - \Phi_2 z^2) = 0$$ lie outside the unit circle. The process has the mean value:

$$\mu = c + \Phi_1 \mu + \Phi_2 \mu$$
$$\mu = \frac{c}{1 - \Phi_1 - \Phi_2}$$
Second order Autoregression

To find the second moments of the AR(2), we write

\[
Y_{k+1} = \mu(1 - \Phi_1 - \Phi_2) + \Phi_1 Y_k + \Phi_2 Y_{k-1} + \varepsilon_k
\]

\[
Y_{k+1} - \mu = \Phi_1(Y_k - \mu) + \Phi_2(Y_{k-1} - \mu) + \varepsilon_k.
\]

Multiplying both sides with \(Y_{k+1-j} - \mu\) we get and taking the expectation yields:

\[
\nu_j = \Phi_1 \nu_{j-1} + \Phi_2 \nu_{j-2} \quad j = 1, 2, 3..
\]

The auto-covariances follow the same second order difference equation as the AR(2) process. The autocorrelation is given by dividing the last equation by \(\nu_0\) we get

\[
\rho_j = \Phi_1 \rho_{j-1} + \Phi_2 \rho_{j-2} \quad j = 1, 2, 3..
\]
Second order Autoregression

For \( j = 1 \) we get for the autocorrelation:

\[
\rho_1 = \Phi_1 + \Phi_2 \rho_1 \]

\[
\rho_1 = \frac{\Phi_1}{1 - \Phi_2},
\]

since \( \rho_0 = 1 \) and \( \rho_{-1} = \rho_1 \). For \( j = 2 \) we get

\[
\rho_2 = \Phi_1 \rho_1 + \Phi_2
\]

\[
\rho_2 = \frac{\Phi_1^2}{1 - \Phi_2} + \Phi_2.
\]

The covariance is computed by

\[
\mathbb{E}[(Y_{k+1} - \mu)(Y_{k+1} - \mu)] = \Phi_1 \mathbb{E}[(Y_k - \mu)(Y_{k+1} - \mu)] + \Phi_2 \mathbb{E}[(Y_{k-1} - \mu)(Y_{k+1} - \mu)] + \mathbb{E}[\varepsilon_k (Y_{k+1} - \mu)],
\]
Second order Autoregression

The covariance equation can be written as

$$\nu_0 = \Phi_1 \nu_1 + \Phi_2 \nu_2 + \sigma^2$$  \( j = 1, 2, 3.. \)

The last term from noticing

$$E[\varepsilon_k (Y_{k+1} - \mu)] = E[(\varepsilon_k)(\Phi_1 (Y_k - \mu)$$

$$+ \Phi_2 (Y_{k-1} - \mu) + \varepsilon_k)]$$

$$= \Phi_1 \nu_0 + \Phi_2 \nu_0 + \sigma^2$$

The covariance equation can be written as

$$\nu_0 = \Phi_1 \rho_1 \nu_0 + \Phi_2 \rho_2 \nu_0 + \sigma^2,$$

$$\nu_0 = \frac{(1 - \Phi_2)\sigma^2}{(1 + \Phi_2)[(1 - \Phi_2^2) - \Phi_1^2]}$$
A p-th order autoregressive process (denotes as AR(p)) has the form:

\[ Y_{k+1} = c + \Phi_1 Y_k + \Phi_2 Y_{k-1} + \ldots + \Phi_p Y_{k-p+1} + \varepsilon_k \]

written in lag notation:

\[ Y_{k+1}(1 - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p) = c + \varepsilon_k \]

The AR(2) is stable if the roots of \((1 - \Phi_1 z - \Phi_2 z^2 - \ldots - \Phi_p z^p) = 0\) lie outside the unit circle. The process has the mean value:

\[
\begin{align*}
E[Y_{k+1}] &= c + \Phi_1 E[Y_k] + \Phi_2 E[Y_{k-1}] + \ldots + \Phi_p E[Y_{k-p+1}] + E[\varepsilon_k] \\
\mu &= c + \Phi_1 \mu + \Phi_2 \mu + \ldots + \Phi_p \mu \\
\mu &= \frac{c}{1 - \Phi_1 - \Phi_2 - \ldots - \Phi_p}
\end{align*}
\]
P-th order Autoregression

The autocovariance is given as

\[ \nu_j = \begin{cases} \Phi_1 \nu_{j-1} + \Phi_2 \nu_{j-2} + \ldots + \Phi_p \nu_{j-p} & j = 1, 2, 3, \ldots \\ \Phi_1 \nu_1 + \Phi_2 \nu_2 + \ldots + \Phi_p \nu_p & j = 0 \end{cases} \]

The equation for the autocorrelation is given as

\[ \rho_j = \Phi_1 \rho_{j-1} + \Phi_2 \rho_{j-2} + \ldots + \Phi_2 \rho_{j-p} \quad j = 1, 2, 3, \ldots \]

which is known as the "Yule-Walker" equation. The equation is again a p-th order difference equation.
Mixed Autoregressive Moving Average Process (ARMA)

A ARMA(p,q) process includes autoregressive elements as well as moving average elements

\[ Y_{k+1} = c + \Phi_1 Y_k + \Phi_2 Y_{k-1} + \ldots + \Phi_p Y_{k-p+1} \]
\[ + \varepsilon_k + \theta_1 \varepsilon_{k-1} + \theta_2 \varepsilon_{k-2} + \ldots + \theta_q \varepsilon_{k-q} . \]

The process is written in Lag form as

\[ Y_{k+1} (1 - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p) = \]
\[ + c + \varepsilon_k (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q) . \]

Provided that the roots of \((1 - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p) = 0\) lie outside the unit circle we obtain the following general form for an ARMA(p,q) process:

\[ Y_{k+1} = \mu + \frac{(1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)}{(1 - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p)} \varepsilon_k . \]
Mixed Autoregressive Moving Average Process (ARMA)

The mean is given as

\[ \mu = \frac{c}{(1 - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p)}. \]

The stationarity of an ARMA process depends only on the autoregressive parameters and not on the moving average parameters. The more convenient form of an ARMA(p,q) process is given as

\[ Y_{k+1} - \mu = \Phi_1 (Y_k - \mu) + \Phi_2 (Y_{k-1} - \mu) + \ldots + \Phi_p (Y_{k-p+1} - \mu) + \varepsilon_k + \theta_1 \varepsilon_{k-1} + \theta_2 \varepsilon_{k-2} + \ldots + \theta_q \varepsilon_{k-q}. \]

The autocovariance equation for \( j > q \) is given as

\[ \nu_j = \Phi_1 \nu_{j-1} + \Phi_2 \nu_{j-2} + \ldots + \Phi_p \nu_{j-p} \quad j = q + 1, q + 2, \ldots \]
Mixed Autoregressive Moving Average Process (ARMA)

Thus after q lags, the autocovariance function $\nu_j$ follows a p-th order difference equation governed by the the autoregressive part. Note that last equation does not hold for $j \leq q$ since the correlation between $\theta_j \varepsilon_{k-j}$ and $Y_{t-j}$ cannot be neglected.

In order to analyze the auto-covariance, we introduce the auto-covariance generating function

$$g_Y(z) = \sum_{j=-\infty}^{\infty} \nu_j z^j$$

which is constructed as taking the j-th auto-covariance and multiplying with $z$ to the power of j and then summing over all possible values of $j$.
Parameter Identification

We start to examine the appropriate log-likelihood function with our framework of the discretized SDEs. Let $\psi \in \mathbb{R}^d$ be the vector of unknown parameters, which belongs to the admissible parameter space $\Psi$, e.g. $a$, $b$, etc. The various system matrices of the discrete models, as well as the variances of stochastic processes, given in the previous section, depend on $\psi$. The likelihood function of the state space model is given by the joint density of the observational data $x = (x_N, x_{N-1}, \ldots, x_1)$

$$l(x, \psi) = p(x_N, x_{N-1}, \ldots, x_1; \psi) \quad (15)$$

which reflects how likely it would have been to have observed the date if $\psi$ were the true values of the parameters.
Parameter Identification

Using the definition of condition probability and employing Bayes’s theorem recursively, we now write the joint density as product of conditional densities

\[
l(x, \psi) = p(x_N|x_{N-1}, \ldots, x_1; \psi) \cdot \ldots \cdot p(x_k|x_{k-1}, \ldots, x_1; \psi) \cdot \ldots \cdot p(x_1; \psi) \tag{16}\n\]

where we approximate the initial density function \(p(x_1; \psi)\) by \(p(x_1|x_0; \psi)\). Furthermore, since we deal with a Markovian system, future values of \(x_l\) with \(l > k\) only depend on \((x_k, x_{k-1}, \ldots, x_1)\) through the current values of \(x_k\), the expression in (16) can be rewritten to depend on only the last observed values, thus it is reduced to

\[
l(x, \psi) = p(x_N|x_{N-1}; \psi) \cdot \ldots \cdot p(x_k|x_{k-1}; \psi) \cdot \ldots \cdot p(x_1; \psi) \tag{17}\n\]
Parameter Identification

A Euler discretized SDE is given as:

\[ x_{k+1} = x_k + f(x_k, k) \Delta t + g(x_k, k) \sqrt{\Delta t} \varepsilon_k \]  

(18)

where \( \varepsilon_k \) is an i.i.d random variable with normal distribution, zero expectation and unit variance. Since \( f \) and \( g \) are deterministic functions, \( x_{k+1} \) is conditional (on \( x_k \) normal distributed with

\[ \begin{align*}
    \mathbb{E}[x_{k+1}] &= x_k + f(x_k, k) \Delta t \\
    \text{Var}[x_{k+1}] &= g(x_k, k) g(x_k, k)^T \Delta t \\
    p(x_{k+1}) &= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\text{Var}[x_{k+1}]|}} e^{-\frac{1}{2} (x_{k+1} - \mathbb{E}[x_{k+1}])^T \text{Var}[x_{k+1}]^{-1} (x_{k+1} - \mathbb{E}[x_{k+1}])} 
\end{align*} \]

(19), (20), (21)
Parameter Identification

The parameter of an SDEs are estimated by maximizing the log-likelihood. To maximize the log-likelihood is equivalent to maximizing the likelihood, since the log function is strictly increasing function. The log likelihood is given as:

\[
\ln(l(x, \psi)) = \ln(p(x_N | x_{N-1}; \psi) \cdot \ldots \cdot p(y_k | x_{k-1}; \psi) \cdot \ldots \cdot p(x_1; \psi)) \tag{22}
\]

\[
\ln(l(x, \psi)) = \sum_{j=1}^{N} \ln(p(x_j | x_{j-1}; \psi) \tag{23}
\]

\[
\ln(l(x, \psi)) = \sum_{j=1}^{N} \left\{ \ln\left((2\pi)^{-\frac{n}{2}} |\text{Var}[x_{k+1}]|^{-\frac{1}{2}}\right) - \frac{1}{2}(x_{k+1} - E[x_{k+1}])^T \text{Var}[x_{k+1}]^{-1}(x_{k+1} - E[x_{k+1}]) \right\} \tag{24}
\]
We want to identify the parameters of a AR(1) process without any constant:

\[ x_{k+1} = \Phi x_k + \sigma \varepsilon_k \]  \hspace{1cm} (25)

which has the log-likelihood function:

\[
\ln(l(x, \psi)) = \sum_{j=1}^{N} \ln((2\pi)^{-\frac{1}{2}}\sigma^{-1}) - \frac{(x_{k+1} - \Phi x_k)^2}{2\sigma^2}.
\]  \hspace{1cm} (26)

The function is maximized by taking the derivatives for \( \Phi \) and \( \sigma \). The log-likelihood function has the same form as an so-called ordinary least-square (OLS) estimation.
Example: Parameter Identification

The solution of the maximum likelihood estimations is:

\[
\hat{\Phi} = \frac{\sum_{j=1}^{N} x_{k+1} x_k}{\sum_{j=1}^{N} x_k^2} \quad (27)
\]

\[
\hat{\sigma} = \frac{1}{N} \sum_{j=1}^{N} (x_{k+1} - \Phi x_k)^2 \quad (28)
\]

The solution for \( \Phi \) is the same solution as for an OLS regression. Furthermore, the solution is exactly the definition of autocorrelation with lag one. The estimation for \( \sigma \) is the empirical standard deviation of the estimation of the empirical white noise, i.e.

\[
\hat{\varepsilon}_k = x_{k+1} - \Phi x_k
\]