Streaming Data Management and Time Series Analysis

Alberto Filosa

28/9/2020

Indice

1 Statistical Prediction **1**
1.1 Optimal Linear Predictor 1
2 Time Series Concepts **2**
2.1 Moving Average Process 3
2.2 Autoregressive Process 3
2.3 Integrated Process 4
2.4 ARIMA Models 4
2.5 ARIMA Regression 4
3 Unobserved Components Model **5**
3.1 Trend 5
3.2 State Space Form 5
3.2.1 Kalman Filter 6
3.3 Stochastic Cycle 7
3.4 Seasonality 7

1 Statistical Prediction

A *Statistical Prediction* is a guess about the value of a random variable Y based on the outcome of other random variables $X_1, \dots, X_m$. A *Predictor* is a function of the random variables: $\hat{Y} = p(X_1, \dots, X_m)$. An optimal prediction is the *Loss Function* which maps the prediction error to it cost; if the prediction error is zero, also the loss is 0 (exact guess = no losses). A loss function can be symmetric about 0 or asymmetric: $l(-x) = l(x)$, but the most used is generally asymmetric, like the *Quadratic loss function*: $l_2(x) = x^2$.

The predictor $\hat{Y} = \hat{p}(X_1, \dots, X_m)$ is optimal for Y with respect to the loss l if minimize the expected loss among the class of measurable functions:

$$\mathbb{E}\ l(Y - \hat{Y}) = \min_{p \in M} \mathbb{E}\ l[Y - p(X_1, \dots, X_m)]$$

The optimal predictor under quadratic loss is the conditional expectation $\hat{Y} = \mathbb{E}[Y|X_1, \dots, X_m]$.

The properties of the conditional expectation are:

- *Linearity*: $\mathbb{E}[aY + bZ + c|X] = a\mathbb{E}[Y|X] + b\mathbb{E}[Z|X] + c$, with a, b, c, constants;
- *Orthogonality of the prediction error*: $\mathbb{E}[(Y - \hat{Y})g(x)] = \mathbb{E}[(Y - \mathbb{E}[Y|X])g(x)] = 0$;
- *Functions of conditioning variables*: $\mathbb{E}[Yg(x)|X] = \mathbb{E}[Y|X]g(x)$;
- *Indipendence with the conditioning variables*: $\mathbb{E}[Y|X] = \mathbb{E}[Y]$ if $X \perp\!\!\!\perp Y$;
- *Law of iterated expecatons*: $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}(Y|X)]$;
- *Law of total variance*: $\mathbb{V}\text{ar}[Y] = \mathbb{E}[\mathbb{V}\text{ar}(Y|X)] + \mathbb{V}\text{ar}[\mathbb{E}(Y|X)]$.

1.1 Optimal Linear Predictor

Sometimes it can be easer to limit the search to a smaller class functions, like linear combination. The main advantage is that the covariance structure of the random variables is all needed to compute the prediction:

$$\mathbb{P}[Y|X_1, \dots, X_m] = \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X)$$

The Properties of the optimal linear predictor are:

1. *Unbiasedness*: $\mathbb{E}[Y - \mathbb{P}(Y|X)] = 0$;
2. ***M**ean **S**quare **E**rror* (**MSE**) of the prediction: $MSE_{lin} = \Sigma_{YY} - \Sigma_{YX}\Sigma_{YY}^{-1}\Sigma_{XY}$
3. *Orthogonality of the prediction error*: $\mathbb{E}[(Y - \mathbb{P}[Y|X])X^t] = 0$;
4. *Linearity*: $\mathbb{P}[aY+bZ+c|X] = a\mathbb{P}[Y|X]+b\mathbb{P}[Z|X]+c$;
5. *Law of iterated projections*: if $\mathbb{E}(X-\mu_x)(Z-\mu_Z)^t = 0$, then $\mathbb{P}[Y|X,Z] = \mu_Y+\mathbb{P}[Y-\mu_Y|X]+\mathbb{P}[Y-\mu_Y|Z]$.

2 Time Series Concepts

A *Time Series* is a sequence of observation ordered to a time index t taking values in an index set S. If S contains finite numbers we speak about of *Discrete* time series y_t, otherwise a *Continuos* time series $y(t)$.

The most important form of time homogeneity is *Stationarity*, defined as time-invariance of the whole probability of the data generating process (*Strict* stationarity) or just of its two moments (*Weak* stationarity).

The process $\{Y_t\}$ is *Strictly* stationary if $\forall k \in \mathbb{N}, h \in \mathbb{Z}$ and $(t_1, \dots, t_k) \in \mathbb{Z}^k$,

$$(Y_{t_1}, \dots, Y_{t_k},) \stackrel{d}{=} (Y_{t_1+h}, \dots, Y_{t_k+h})$$

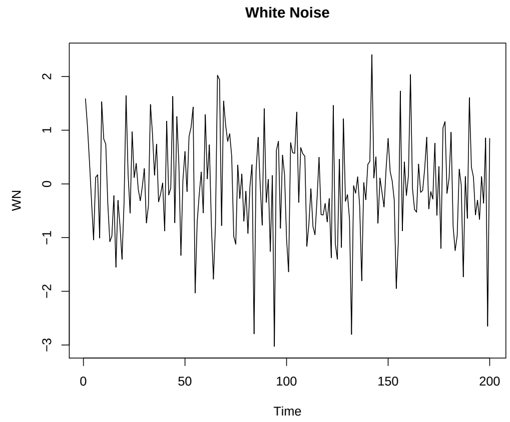
The process $\{Y_t\}$ is *Weakly* stationary if, $\forall h, t \in \mathbb{Z}$, with $\gamma(0) < \infty$

$$\mathbb{E}(Y_t) = \mu$$
$$\mathbb{C}\mathrm{ov}(Y_t, Y_{t-h}) = \gamma(h)$$

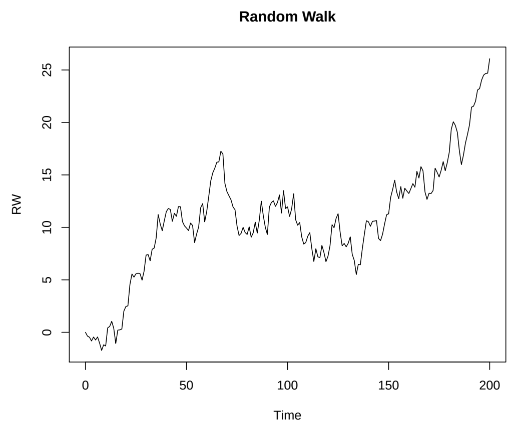
If a time series is strictly stationary, then it is also weakly stationary if and only if $\mathbb{V}\mathrm{ar}(Y_t) < \infty$. If a time series is a Gaussian process, then strict and weak stationarity are equivalent.

The most elementary stationarity process is the White Noise, a time series that show no autocorrelation. A stochastic process is *White Noise* if has $\mu = 0, \sigma^2 > 0$ and covariance function

$$\gamma(h) = \begin{cases} \sigma^2 & \text{for } h = 0 \\ 0 & \text{for } h \neq 0 \end{cases}$$



If a time series is non-stationary, then the process is called as *Random Walk*. It is a cumulative sum of a zero mean White Noise series:



The *Autocovariance Function* is a function characterized by a weakly stationary process, while the *Autocorrellation Function* (**ACF**) is the scale independent version of the autocovariance function. It measures the linear relationship between lagged values of a time series.

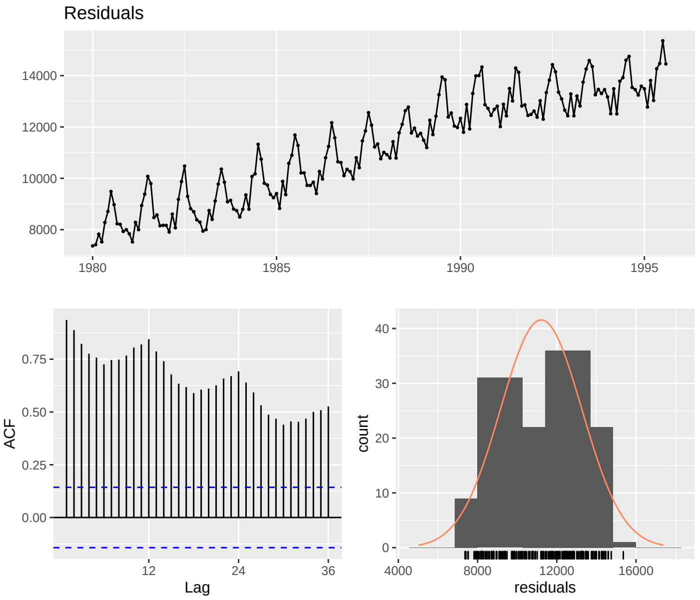
If $\{Y_t\}$ is a stationary process with autocovariance $\gamma(\cdot)$, then its ACF is $p(h) = \mathbb{C}\mathrm{or}(Y_t, Y_{t-h}) = \gamma(h)/\gamma(0)$.

An alternative summary of the linear dependence of a stationary process can be obtained from the *Partial Autocorrellation Function*, that measures the correlation between Y_t and Y_{t-h} after their linear dependence on the interventing random variables has been removed:

$$\alpha(h) = \mathbb{C}or[Y_t - \mathbb{P}(Y_t|Y_{t-1|t-h+1}), Y_{t-h} - \mathbb{P}(Y_{t-h}|Y_{t-1|t-h+1})]$$

When data have a *Trend*, autocorrelations for small lags tend to be large and positive because observation nearby in time are also nearby in size, so the ACF tend to have positive values that slowly decrease as the lags increase.

When data are seasonal, autocorrelations will be larger for the seasonal lags than for other lags.



The slow decrease in ACF as the lags increase is due to the trend, while the scalloped shape is due to seasonality.

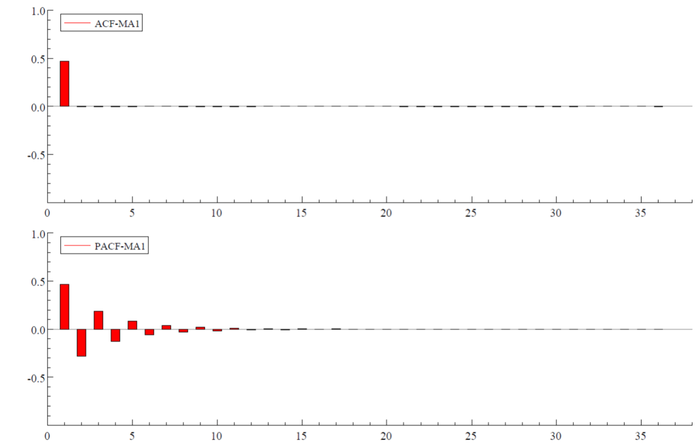
2.1 Moving Average Process

A ***M***oving ***A***verage Process (**MA**) is a process that estimates the trend-cycle at time t obtained by averaging values of the time series within q periods of t. The average eliminates some of the randomness in the data, leaving a smooth trend-cycle component.

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$
$$Y_{t-1} = \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q-1}$$

A $MA(q)$ process has $PACF = 0$ for $h > p$ and its characteristic equation $1 + \theta_1 X + \cdots + \theta_q X^q = 0$ that has only external solution of the unitary circle is a process such that $Y_t = k + \psi_1 Y_{t-1} + \cdots + \psi_q Y_{t-q} + \varepsilon_t$.

A property of $MA(q)$ models in general is that there are nonzero autocorrelations for the first q lags and autocorrelations $= 0$ for all lags $> q$. The theoretical PACF does not shut off, but instead tapers toward 0 in some manner. A clearer pattern for an $MA(q)$ model is in the ACF. The ACF will have non-zero autocorrelations only at lags involved in the model.



2.2 Autoregressive Process

An Autoregressive Process (**AR**) is a process that forecasts the Y variable using a linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself.

$$Y_t = k + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

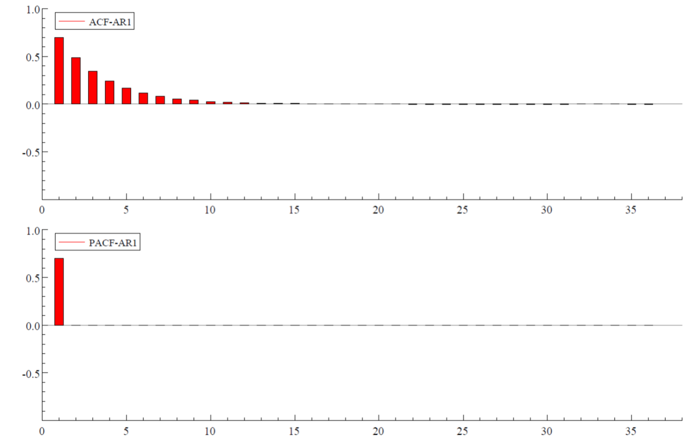
An $AR(p)$ process is stationary for the characteristic equation $1 - \phi_1 X - \cdots - \phi_p X^p = 0$ and all solutions are external of the unitary circle

If Y_t is stationary, $\mathbb{E}(Y_t) = \frac{k}{1-\phi_1-\cdots-\phi_p}$.

The $AR(1)$ process is $Y_t = k + \phi Y_{t-1} + \varepsilon_1$, so $1 - \phi x = 0 \implies x = 1/\phi$. $AR(1)$ is stationary if $|\phi| < 1$. If in $AR(1)$ $\phi = 1$ we obtain a non stationary process (integrate) called Random Walk:

$$Y_t = k + Y_{t-1} + \varepsilon$$

For a positive value of ϕ_p, the ACF exponentially decreases to 0 as the lag increases. For a negative value, the ACF also exponentially decays to 0 as the lag increases, but the algebric signs for the autocorrelations alternate between positive and negative. The theoretical PACF *shuts off* (close to 0) past the order of the model.



2.3 Integrated Process

An *Integrated Process* $\{Y_t\} \sim I(d)$ is a non stationary process, but its first difference is stationary:

$$\Delta Z_t = Z_t - Z_{t-1} \sim I(0)$$

A Z_t process is integrated of order d if it is not stationary, $\Delta^{d-1} Z_{t-1}$ non stationary, while $\Delta^d Z_t$ is stationary.

The process $\{Y_t\}$ is ARMA(p, q) if it is stationary and satisfies:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1 Z_{t-1} + \cdots + \theta_1 Y_{t-q}$$

An $AR(p)$ process is integrated in order d if there are d solution $x = 1$ and the other $|x| > 1|$.

2.4 ARIMA Models

If we combine differencing with autoregression and a moving average model, we obtain a non-seasonal ***A***uto***R***egressive ***I***ntegrated ***M***oving ***A***verage (**ARIMA**).

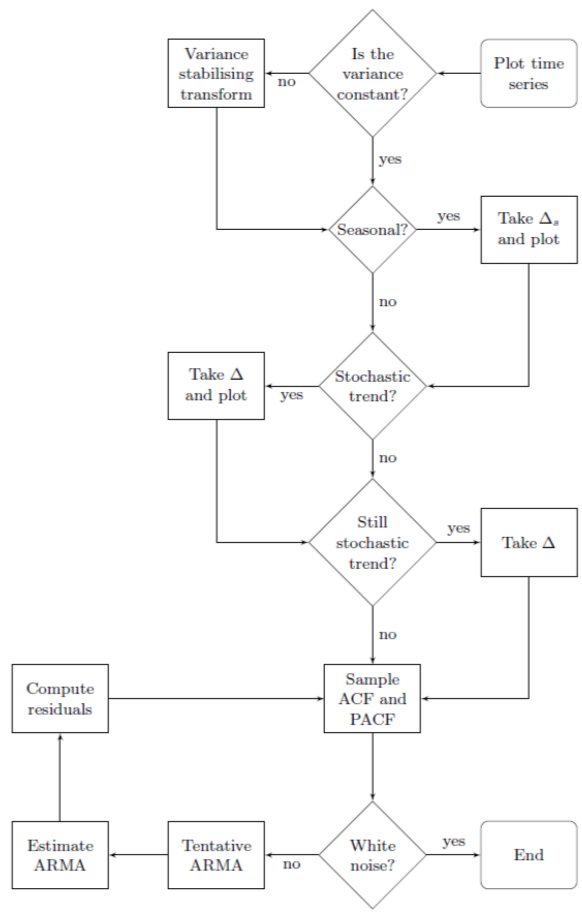
$$Y'_t = c + \phi_1 Y'_{t-1} + \cdots + \phi_p Y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where Y'_t is the differenced series. This is an $ARIMA(p, d, q)$, where p is the order of the ***A***uto***R***egressive part, d is the degree of first differencing involved (***I***ntegrated) and q is the order of the ***M***oving ***A***verage part.

The backward shift operator $B(y_t) = y_{t-1}$ is a useful notational device when working with time series lags, so it has the effect of shifting the data back one period. The backward shift operator is convenient for describing the process of differencing. A first difference can be written as:

$$y'_t = y_t - y_{t-1} = y_t - B(y_t) = (1 - B)y_t$$

Backshift notation is particularly useful when combining differences, as the operator can be treated using ordinary algebraic rules. In particular, terms involving B can be multiplied together.



2.5 ARIMA Regression

The time series models allow for the inclusion of information from past observations of a series, but not for the inclusion of other information that may also be relevant:

$$\Delta^d y_t = \beta_t \Delta^d X_t + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t$$

$$\Delta^d (y_t - \beta_t X_t) = \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t$$

$$\phi(B) \Delta^d (y_t - \beta_t X_t) = \theta(B) \varepsilon_t$$

The data may follow an $ARIMA(p, d, 0)$ model if the ACF is exponentially decaying or sinusoidal and there is a significant spike at lag p in PACF but none beyond lag p.

The data may follow an $ARIMA(0, d, q)$ model if the there is a significant spike at lag q in ACF but none beyond lag q and PACF is exponentially decaying or sinusoidal.

3 Unobserved Components Model

A natural way humans tend to think about time series is as sum of non directly observable components such as trends, seasonality and a component cycle. ***U***nobserved ***C***omoponents ***M***odels (**UCM**) select the best features of stochastic framework as ARIMA models, but also they tend to perform very well in forecasting.

The observable time series of UCM is the sum of Trend (μ_t), Cycle (ψ_t), Seasonality (γ_t) and White Noise (ε_t):

$$Y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t$$

Some of these components could be skipped and some other could be added. Also, they can be seen as stochastic version of the deterministic functions of time. In the next sections we see how to build stochastically evolving components.

3.1 Trend

The *Trend* is responsible for the mean variation process, usually adopted in UCM as the ***L***ocal ***L***inear ***T***rend (**LLT**). Let us take a linear function defined as follows:

$$\mu_t = \mu_0 + \beta_0 t$$

where μ_0 is the intercept and β_0 the slope and write it in incremental form:

$$\mu_t = \mu_{t-1} + \beta_0$$

By adding the White Noise η_n, we obtain a Random Walk with Drift. In this case we can interpreted μ_t as a linear trend with a Random Walk intercept, but the slope remains unchanged.

It is possible obtain a time-varying slope in trends, easily achieved letting the slope evolves as a Random Walk:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_0 + \eta_n \\ \beta_t &= \beta_{t-1} + \xi_t \end{aligned}$$

These equations define the local linear trend interpreted as a linear trend where both intercept and slope evolve in time as Random Walks.

The local linear trend has different special case of interest obtained by fixing the value of the variances:

- *Deterministic Linear Trend* if $\sigma^2_\eta = \sigma^2_\xi = 0$;
- *Random Walk with Drift* β_0 if $\sigma^2_\xi = 0$ (slope constant);
- *Random Walk* if $\sigma^2_\xi = \beta_0 = 0$ (slope = 0);
- *Integrated Random Walk* if $\sigma^2_\eta = 0$, with a very smooth trend.

Obs.: the local linear trend can be also seen as an ARIMA process: if $\sigma^2_\xi > 0$, $\mu_t \sim I(2)$ process as trend is non-stationary, while its second difference is stationary.

3.2 State Space Form

The state space form is a system of equation in which one or more observable time series are linearly related to a set of unobservable state variables. It is defined by the following system of equation:

$$\begin{cases} Y_t &= d_t + Z_t\alpha_t + \varepsilon_t \quad \text{Observation Equation} \\ \alpha_{t+1} &= c_t + T_t\alpha_t + R_t\eta_t \quad \text{State Equation} \end{cases}$$

where $\varepsilon_t \sim WN(0, H_t)$, $\eta_t \sim WN(0, Q_t)$, uncorrelated random variables with zero mean and covariance matrix respectably H_t and Q_t.

The initialization is the process where the state vector $a_{1|0} = \mathbb{E}(\alpha_1)$ and $P_{1|0} = \mathbb{E}(\alpha_1 - a_{1|0})(\alpha_1 - a_{1|0})^t$.

Example. Let's transform a Local Linear Trend with White Noise in the State Space form. The LLT has this system of equation:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_0 + \eta_n \\ \beta_t &= \beta_{t-1} + \xi_t \end{aligned}$$

with noise $y_t = \mu_t + \varepsilon_t$. The state space form of the LLT is the following:

$$\begin{cases} \alpha_{t+1} = T\alpha_t + R\eta_t = \begin{bmatrix} \mu_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix} \quad \text{State I} \\ Y_t = Z\alpha_t + \varepsilon_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \varepsilon_t \quad \text{Observation Equation} \end{cases}$$

where $\eta_t = \begin{bmatrix} \eta_t \\ \zeta^2 \end{bmatrix}$ and $Q = \begin{bmatrix} \sigma^2_\eta & 0 \\ 0 & \sigma^2_\zeta \end{bmatrix}$.

The initialization process is the following:

$$a_{1|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad P_{1|0} = \begin{bmatrix} \infty & 0 \\ 0 & \infty \end{bmatrix}$$

Obs. The Integrated Random Walk is a LLT with the variance of the first disturbance set to zero.

3.2.1 Kalman Filter

Suppose that all parameters in the state space form are known; the only unknown quantities are unobservable components, specified as random variables, then the inference to carry out is the statistical prediction:

- Forecasting, the SP of α_t based on Y_s $(s < t)$;
- Filtering, SP of α_t based on Y_t;
- Smoothing, SP of α_t based on Y_s $(s > t)$.

It is possible to build the optimal linear predictor that is better than any other predictor assuming if ε_t, $v_t \sim WN$ and the initial state α_1 are jointly Gaussian.

Consider the following notation:

$$\begin{aligned} a_{t|s} &= \mathbb{P}[\alpha_t|Y_t] \\ P_{t|s} &= \mathbb{E}[\alpha_t - a_{t|s}][\alpha_t - a_{t|s}]^t \end{aligned}$$

The *Kalman Filter* is an algorithm for computing the pair $\{a_{t|t-1}, P_{t|t-1}\}$ starting from $\{a_{t-1|t-1}, P_{t-1|t-1}\}$ and $\{a_{t|t}, P_{t|t}\}$ starting from $\{a_{t|t-1}, P_{t|t-1}\}$. Perciò proietta le stime di y_t partendo basate sulle osservazioni precedenti.

It also provides the sequence of innovations with the relative covariance matrix, used to compute the Gaussian likelihood of the model in state space form.

- Prediction Step:

$$\begin{cases} a_{t|t-1} &= \mathbb{P}(\alpha_t|y_1, \dots, y_{t-1}) \\ P_{t|t-1} &= \mathbb{E}[\alpha_t - a_{t|t-1}][\alpha_t - a_{t|t-1}]^t \end{cases}$$

- One-Step-Ahead Forecast and Innovation:

$$\begin{cases} \hat{y}_{t|t-1} &= \mathbb{P}(\alpha_t|y_1, \dots, y_{t-1}) \\ i_t &= y_t - \hat{y}_{t|t-1} \\ F_t &= \mathbb{E}[i_t i_t^t] = \mathbb{E}[y_t - \hat{y}_{t|t-1}][y_t - \hat{y}_{t|t-1}]^t \end{cases}$$

- Updating Step:

$$\begin{cases} a_{t|t} &= \mathbb{P}(\alpha_t|y_1, \dots, y_t) \\ P_{t|t} &= \mathbb{E}[\alpha_t - a_{t|t}][\alpha_t - a_{t|t}]^t \end{cases}$$

$$(a_{1|0}, P_{1|0}) \rightarrow (a_{1|1}, P_{1|1}) \rightarrow (a_{2|1}, P_{2|1}) \rightarrow \dots (a_{n|n}, P_{n|n})$$

The linear prediction of the state vector α_t based on $y_s = \{Y_1, \dots, Y_s\}$ is called *Smoothing*. The fixed interval smoother provides prediction for the state vector based on the whole time series:

$$a_{t|n} = \mathbb{P}[\alpha_t|Y_n]$$

$$\begin{aligned} \hat{y}_{t|t-1} &= \mathbb{P}(y_t|y_{t-1}) \\ \hat{y}_{t|n} &= \mathbb{P}(y_t|y_n) \end{aligned}$$

The Disturbance Smoothing computes the prediction of the White Noise sequences based on the whole time series, useful for identifying outliers:

$$\begin{aligned} \hat{\varepsilon}_{t|n} &= \mathbb{P}(\varepsilon_t|y_n) \\ V^{\varepsilon}_{t|n} &= \mathbb{E}[\varepsilon_t - \hat{\varepsilon}_{t|n}][\varepsilon_t - \hat{\varepsilon}_{t|n}]^t \\ U^{\varepsilon}_{t|n} &= \mathbb{E}[\hat{\varepsilon}_{t|n}]\hat{\varepsilon}_{t|n}]^t] \\ \hat{\eta}_{t|n} &= \mathbb{P}(\eta_n|y_n) \\ V^{\varepsilon}_{t|n} &= \mathbb{E}[(\eta_n - \hat{\eta}_{t|n})(\eta_n - \hat{\eta}_{t|n})^t] \\ U^{\varepsilon}_{t|n} &= \mathbb{E}[\hat{\eta}_{t|n}\hat{\eta}_{t|n}]^t] \end{aligned}$$

If observation are Normally Distributed, the Kalman Filter allows the construction of the likelihood function. Using the definition of conditional density, it is possible to factorise the joint density of the data:

$$f_\theta(y_1, \dots, y_n) = f_\theta(y_1) f_\theta(y_2|y_1) \dots f_\theta(y_n|y_{n-1}, \dots, y_1)$$

Under this assumption, the conditional distribution is also Normal and the Kalman Filter provides the conditional mean and variance:

$$\mathbb{E}(Y_t|Y_1, dots, Y_{t-1}) = \hat{y}_{t|t-1} \mathbb{V}ar(Y_t|Y_1, dots, Y_{t-1}) = F_t$$

The likelihood is the joint density of the sample path of the unknown parameters θ:

$$L(\theta) = \prod_{t=1}^{n} f_\theta(y_t|y_{t-1}, \dots, y_1)$$

It is generally better to work on the log-likelihood function:

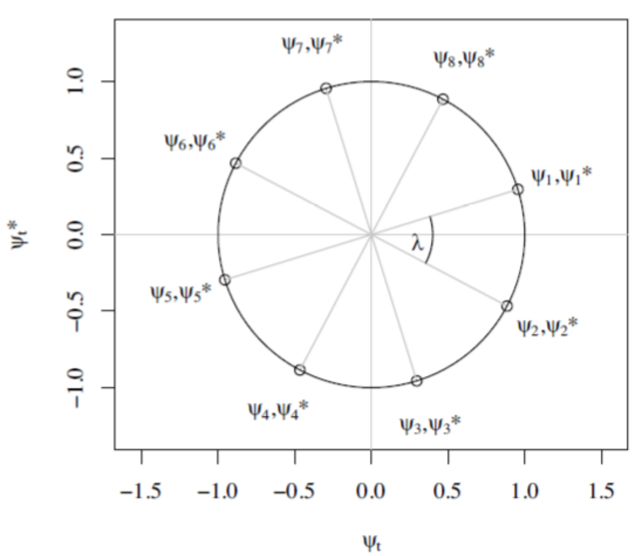
$$\begin{aligned} l(\theta) = \log f(y_1, \dots, y_n) = \sum_{t=1}^{n} \log f(y_t|y_1, \dots, y_{t-1}) = \\ = \sum_{t=1}^{n} -1/2(\log|F_t(\theta)| + \\ = (y_t - \hat{y}_{t|t-1}(\theta)^t F_t(\theta)^{-1}(y_t - \hat{y}_{t|t-1}(\theta)^t \\ \implies \hat{\theta}_n = \max l(\theta) \end{aligned}$$

3.3 Stochastic Cycle

A cycle occurs when the data exhibit rises and falls that are not of a fixed frequency. The natural deterministic function for generating a *Cycle* with frequency λ is the sinusoid $R\cos(\lambda t + \phi)$ where R is the amplitude and ϕ the phase. A geometrical way to generate a sinusoid is taking a circumference of radius equal to R and letting a point move on it. The ψ_t cycle is stationary:

$$\begin{bmatrix} \psi_{t+1} \\ \psi^*_{t+1} \end{bmatrix} = \rho \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ -\sin(\lambda) & \cos(\lambda) \end{bmatrix} \begin{bmatrix} \psi_t \\ \psi^*_t \end{bmatrix} + \begin{bmatrix} k_t \\ k^*_t \end{bmatrix}$$

where $\rho \in [0, 1]$ is the damping factor, $\lambda \in [0, \pi]$ the frequency of the cycle and k_t, k^*_t are indipendent WN sequences with common variance σ^2_k.



There is a unique casual and stationary solution with $\mathbb{E}[\psi_t] = 0$, $\mathbb{E}[\psi_{t+h}\psi^t_t] = \frac{\sigma^2_k}{1-\rho^2} I^2$ and $\psi_t \sim ARMA(2,1)$ process with complex roots in the AR polynomial.

When $\rho = 1$ and $\mathbb{E}[\psi_t] = 0$, the cycle is not stationary, while if $\rho < 1$ the cycle is stationary.

3.4 Seasonality

A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality always regards a fixed and known frequency. The seasonal component is generally modelled using the stochastic dummy form and the trigonometric one.

Consider a time series with seasonality s:

$$\sum_{j=1}^{s/2} a_j \cos(\frac{2\pi}{s} jt) + b_j \sin(\frac{2\pi}{s} jt)$$

The *Trigonometric Form* is given by $\gamma_t = \sum_{j=1}^{s/2} \gamma^{(j)}_t$, $\gamma^{(j)}_t$ is the nonstationarity stochastic cycle:

$$\begin{bmatrix} \gamma^{(j)}_{t+1} \\ \gamma^{*(j)}_{t+1} \end{bmatrix} = \begin{bmatrix} \cos(\frac{2j\pi}{s}) & \sin(\frac{2j\pi}{s}) \\ -\sin(\frac{2j\pi}{s}) & \cos(\frac{2j\pi}{s}) \end{bmatrix} \begin{bmatrix} \gamma^{(j)}_t \\ \gamma^{*(j)}_t \end{bmatrix} + \begin{bmatrix} w^{(j)}_t \\ w^{*(j)}_t \end{bmatrix}$$

The arguments of the sinusoids are called seasonal frequencies. When s is even and $j = s/2$, the second equation of the previous system can be omitted because $\sin(\pi) = 0$ and the first equation is reduced as $\gamma^{(s/2)}_{t+1} = -\gamma^{(s/2)}_t - w^{(s/2)}_t$ which does not depend on the value of the second element.

An alternative way (dummy form) of modeling seasonal component is defining s variables that evolve as RW. Let γ_t represent the effect of the season a time t,

$$\gamma_t = -\sum_{s=1}^{s-1} \gamma_{t-i} + w_t$$

where $w_t \sim WN(0, \sigma^2_w)$. The most straightforward way to make this component evolve stochastically over time is by adding mean-zero random shocks.