Streaming Data Management and Time Series Analysis

Alberto Filosa

28/9/2020

Indice

1 Previsione Statistica

Una Previsione Statistica è una stima del valore della variable casuale Y basato sull'esito di altre variabili casuali indipendenti $X_1, \dots, X_m$. Uno stimatore è una funzione delle variabile casuale: $\hat{Y} = p(X_1, \dots, X_m)$. uno stimatore ottimale è la funzione di perdita che mappa le previsioni degli errori sui costi; se la previsione di errore è nulla, lo è anche la perdita (no perdita di informazione). Una funzione di perdita può essere simmetrica o asimmetrica: $l(-x) = l(x)$; generalmente viene utilizzata asimmetrica, come la funzione di perdita quadratica $l_2(x) = x^2$.

Lo stimatore $\hat{Y} = \hat{p}(X_1, \dots, X_m)$ è ottimale se minimizza la perdita attesa tra le funzioni di classe misurabili:

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

Lo stimatore ottimale sollo la funzione di perdita quadratica è la previsione attesa condizionata: $\hat{Y} = \mathbb{E}[Y|X_1, \dots, X_m]$.

Le proprietà della previsione attesa condizionata sono:

- *Linearità*: $\mathbb{E}[aY+bZ+c|X] = a\mathbb{E}[Y|X]+b\mathbb{E}[Z|X]+c$, con a, b, c, costanti;
- *Ortogonalità*: $\mathbb{E}[(Y - \hat{Y})g(x)] = \mathbb{E}[(Y - \mathbb{E}[Y|X])g(x)] = 0$;
- *Funzioni Condizionate*: $\mathbb{E}[Yg(x)|X] = \mathbb{E}[Y|X]g(x)$;
- *Indipendenza del Valore Atteso Condizionato*: $\mathbb{E}[Y|X] = \mathbb{E}[Y]$ if $X \perp\!\!\!\perp Y$;
- *Legge dei Valori Attesi Iterati*: $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}(Y|X)]$;
- *Legge della Varianza Totale*: $\mathbb{V}\text{ar}[Y] = \mathbb{E}[\mathbb{V}\text{ar}(Y|X)] + \mathbb{V}\text{ar}[\mathbb{E}(Y|X)]$.

1.1 Stimatore Ottimale Lineare

Alcune volte è più semplice limitare la ricerca con funzioni semplici, ad esempio le combinazioni lineari. Il vantaggio principale è che la struttura della covarianza delle v.c. è tutto ciò che serve per la previsione:

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

Le proprietà dell stimatore ottimo lineare sono:

1. *Unbiasedness*: $\mathbb{E}[Y - \mathbb{P}(Y|X)] = 0$;
2. ***M**ean **S**quare **E**rror* (**MSE**) delle previsioni: $MSE_{lin} = \Sigma_{YY} - \Sigma_{YX}\Sigma_{YY}^{-1}\Sigma_{XY}$
3. *Ortogonalità*: $\mathbb{E}[(Y - \mathbb{P}[Y|X])X^t] = 0$;
4. *Linearità*: $\mathbb{P}[aY+bZ+c|X] = a\mathbb{P}[Y|X]+b\mathbb{P}[Z|X]+c$;
5. *Legge delle Proiezioni Iterate*: se $\mathbb{E}(X - \mu_x)(Z - \mu_Z)^t = 0$, allora $\mathbb{P}[Y|X,Z] = \mu_Y + \mathbb{P}[Y - \mu_Y|X] + \mathbb{P}[Y - \mu_Y|Z]$.

1.2 Serie Storica

Una Serie Storica è una sequenza di osservazioni ordinate in un temp t che prende valori con un indice s. Se s contiene osservazioni finite si parla di una serie storica *Discreta* y_t, altrimenti è una serie storica *Continua* $y(t)$.

Il concetto più importante di una serie storica è la *Stazionarietà*, definita come l'invarianza nel tempo dell'intero processo dei dati (stazionarietà *Forte*) oppure nei primi due momenti (stazionarietà *Debole*).

In particolare, un processo $\{Y_t\}$ è detto stazionario in senso stretto se $\forall k \in \mathbb{N}, h \in \mathbb{Z}$ e $(t_1, \dots, t_k) \in \mathbb{Z}^k$,

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

Un processo $\{Y_t\}$ è detto stazionario in senso debole se $\forall h, t \in \mathbb{Z}$, with $\gamma(0) < \infty$

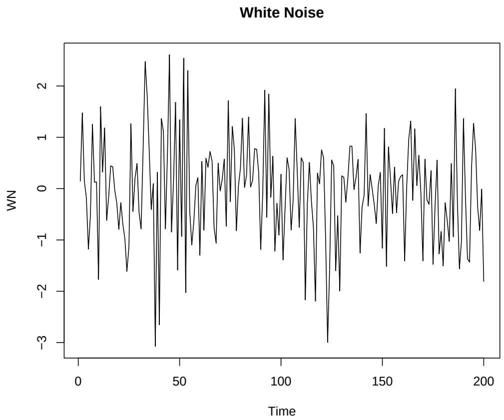
$$\mathbb{E}(Y_t) = \mu$$
$$\mathbb{C}\mathrm{ov}(Y_t, Y_{t-h}) = \gamma(h)$$

ovvero se media e varianza non dipendono dal tempo e la covarianza del processo con se stesso anticipato, detta funzione di auto-covarianza, dipende solo dalla distanza temporale.

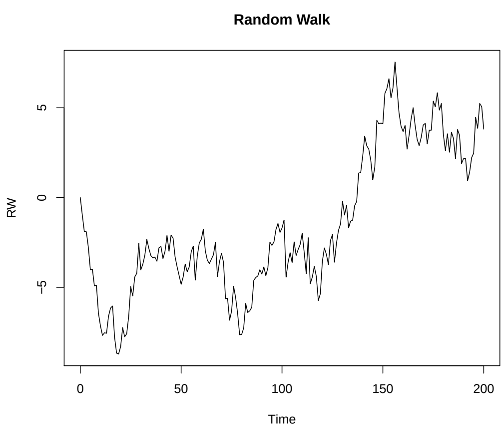
Se una serie storica è stazionaria in senso stretto, allora lo è anche in senso debole se e solo se $\mathbb{V}\mathrm{ar}(Y_t) < \infty$. Se una serie storica è distribuito come una Gaussiana, allora la stazionarietà stretta e debole coincidono.

Il processo stazionario più semplice è chiamato *White Noise*, una serie storica senza autocorrelazione. Un processo stocastico è definito White Noise se $\mu = 0, \sigma^2 > 0$ e la sua funzione di covarianza pari a:

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



Se la serie storica è un processo non stazionario, allora viene chiamato come *Random Walk*, definito come una somma cumulativa di $WN(0, \sigma^2)$:



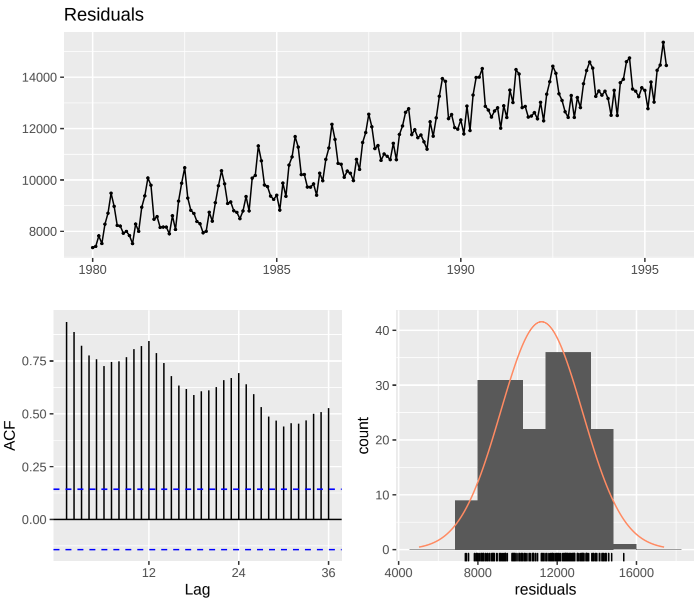
La funzione di *Autocovarianza* è una funzione caratterizzata da un processo stazionario debole, mentre la funzione di *Autocorrelazione* (ACF) è la versione di scala indipendente della autocovarianza. Essa misura la relazione lineare tra valori ritardati di una serie storica.

Se $\{Y_t\}$ è un processo stazionario con autocovarianza $\gamma(\cdot)$, allora la sua ACF è pari a $p(h) = \mathbb{C}\mathrm{or}(Y_t, Y_{t-h}) = \gamma(h)/\gamma(0)$.

Un altro modo per osservare la dipendenza lineare di un processo stocastico è con la funzione di *Autocorrelazione Parziale* (PACF), che misura la correlaione tra Y_t e Y_{t-h} dopo che la loro dipendenza lineare sulle variabili casuali è stata rimossa:

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

Quando i dati hanno un *Trend*, la autocorrelazione tende ad avere alti valori per piccoli ritardi in quanto le osservazioni vicine in termini di tempo lo sono anche in grandezza; in particolare, ACF ha valori positivi che lentamente diminuiscono a 0. Quando i dati sono *Stagionali*, l'autocorrelazione sarà più grande per i ritardi stagionali che per gli altri.



L'andamento decrescente nel grafico ACF all'aumento del ritardo è dovuto al trend, mentre l'andamento a creste è dovuto alla stagionalità.

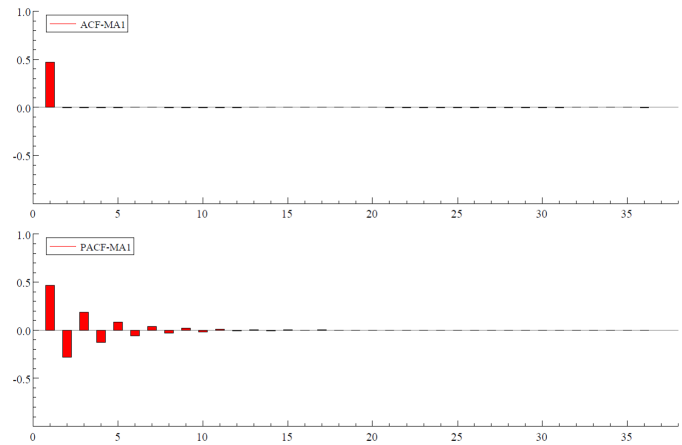
1.3 Processo a Media Mobile

Un processo a *Media Mobile*, in inglese ***M**oving **A**verage* Process (**MA**), è un processo che stima il ciclo di trend dei dati al tempo t ottenuto come la media dei valori della serie storica all'interno di q periodi di t. La media è utile per eliminare una parte della casualità dei dati, lasciando una componente ciclica del trend:

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

Un processo $MA(q)$ ha la funzione di autocorrelazione che si annulla per $h > q$. Inoltre, un processo $MA(q)$ per cui l'equazione caratteristica $1 + \theta_1 X + \dots + \theta_q X^q = 0$ ha solo soluzioni esterne al cerchio unitario ($|x > 1|$ se x è la soluzione della equazione) ed è rappresentabile come $Y_t = k + \psi_1 Y_{t-1} + \dots + \psi_q Y_{t-q} + \varepsilon_t$.

Una particolare proprietà di un processo $MA(q)$ è che in generale le autocorrelazioni per i primi q ritardi sono non nulli e nulli per tutti ritardi maggiori di q. Un pattern più chiaro per identificare un processo $MA(q)$ è osservare il grafico ACF, nella quale ha valori non nulli solamente per i ritardi coinvolti nel modello.



1.4 Processo Autoregressivo

Un processo *Autoregressivo*, in inglese Autoregressive Process (**AR**), è un processo nella quale si vuole prevedere la variabile Y come combinazione lineare dei valori passati. Il termine autoregressivo indica a regressione della variabile con sè stessa:

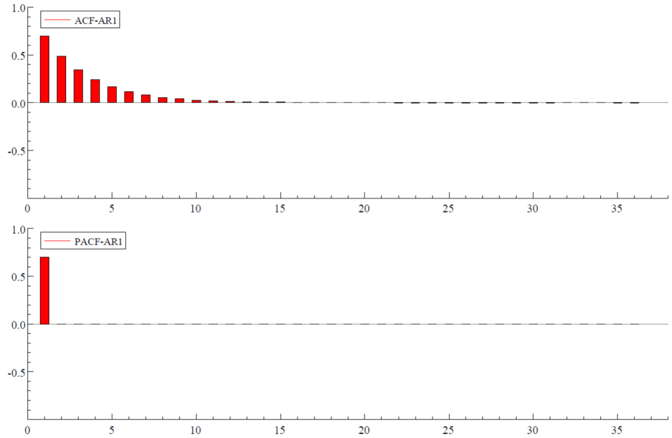
$$Y_t = k + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

Un processo $AR(p)$ è stazionario se nella equazione caratteristica $1 - \phi_1 X - \dots - \phi_p X^p = 0$ tutte le soluzioni sono esterne al cerchio unitario ($|x > 1|$ se x è la soluzione della equazione). Se Y_t è stazionario allora la sua media sarà $\mathbb{E}(Y_t) = \frac{k}{1-\phi_1-\dots-\phi_p}$.

Il processo $AR(1)$ è definito come $Y_t = k + \phi Y_{t-1} + \varepsilon_1$, di conseguenza $1 - \phi x = 0 \implies x = 1/\phi$. Esso è stazionario se $|\phi| < 1$; se in $AR(1)$ $\phi = 1$ si ottiene un processo stazionario (integrato) chiamato Random Walk:

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

Per valori positivi di ϕ_p, il grafico ACF diminuisce a 0 all'aumentare del ritardo. Per valori negativi, invece, il grafico ACF diminuisce anch'essa a 0, ma alterna valori positivi e negativi. Il grafico PACF si spegne dopo aver superato l'ordine p del modello.



1.5 Processo Integrato

Un *Processo Integrato* $\{Y_t\} \sim I(d)$ è un processo non stazionario nella quale la differenza prima è stazionaria. Ad esempio, si consideri un processo integrato di ordine 1:

$$\begin{aligned} Y_t &= Y_{t-1} + \varepsilon_t \sim I(0) \\ Y_t - Y_{t-1} &= \varepsilon_t \sim I(1) \end{aligned}$$

un processo Z_t è integrato di ordine d se Z_t non è stazionario, $\Delta^{d-1} Z_{t-1}$, mentre $\Delta^d Z_t$ è stazionario.

In un processo $AR(p)$, $\phi_p(B)Y_t = \varepsilon_t$, se esistono d soluzioni con $x = 1$ e le restanti $|x| > 1$, esso è un processo stazionario di ordine d:

$$\Delta^d Y_t = AR(p-d)$$

Un processo $\{Y_t\}$ è chiamato ARMA(p,q) se è stazionario e soddisfa la seguente equazione:

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

1.6 Modello ARIMA

Se si combinano i modelli autoregressivi ed a media mobile, si otteine un modello non stagionale ***A***uto***R***egressive ***I***ntegrated ***M***oving ***A***verage (**ARIMA**).

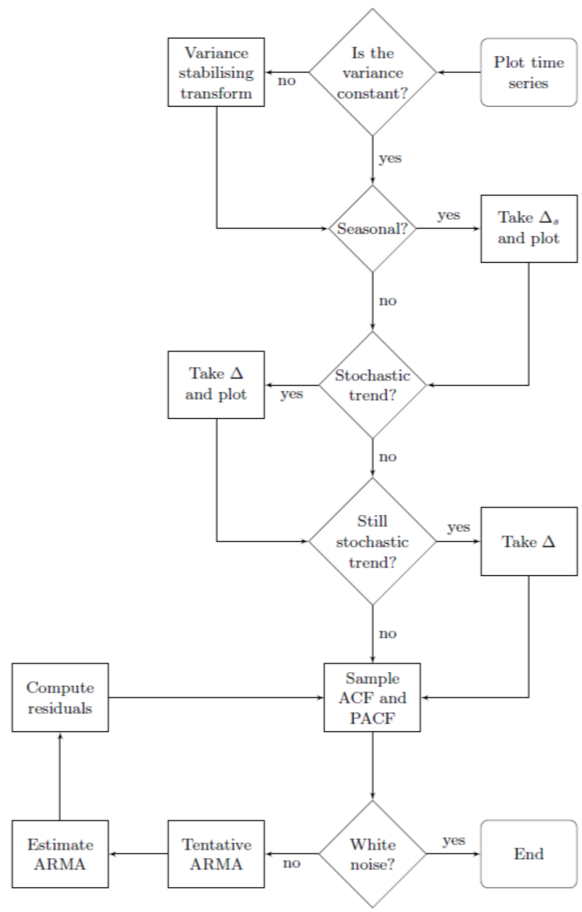
$$Y'_t = c + \phi_1 Y'_{t-1} + \cdots + \phi_p Y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

con Y'_t le serie differenziate. Il modello $ARIMA(p,d,q)$ presenta la parte p AutoRegressiva, il primo grado di Integrazione d e la parte q della Media Mobile.

L'operatore ritardo $B(y_t) = y_{t-1}$ è una notazione molto utile per snellire l'equazione e sposta di un periodo precedente la serie storica. La prima differenze è possibile scriverla come segue:

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

L'operatore ritardo è utile per combinare le differenze, tanto che è possibile utilizzare semplici regole algebriche. In particolare, è possibile coinvolgere più operatori ritardo B insieme.



1.7 Regressione ARIMA

I modelli di serie storica permettono di osservare le informazioni dei periodi presenti, ma non di altre informazioni che possono essere rilevanti:

$$\begin{aligned} \Delta^d y_t &= \beta_t \Delta^d X_t + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t \\ \Delta^d (y_t - \beta_t X_t) &= \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t \\ \phi(B) \Delta^d (y_t - \beta_t X_t) &= \theta(B) \varepsilon_t \end{aligned}$$

I modelli $ARIMA(p, d, 0)$ hanno un grafico ACF che decresce esponenzialmente o in modo sinusoidale ed il grafico PACF presenta valori molto alti fino al ritardo p, mentre valori prossimi a 0 superato il ritardo p.

I modelli $ARIMA(0, d, q)$ hanno un valore molto alto fino al ritardo q nel grafico ACF e valori prossimi a 0 superato il ritardo, mentre el grafico PACF il valore decresce esponenzialmente o in in termini sinusoidali.

2 Modelli a Componenti non Osservabili

Una serie storica può essere considerata come somma di componenti non osservabili, come trend, stagionalità e componente ciclica. I Modelli a Componenti non Osservabili, in inglese ***U***nobserved ***C***omoponents ***M***odels (**UCM**), selezionano le componenti stocastiche migliori dei modelli ARIMA, ma sono anche molto utili nelle previsioni.

La serie storica UCM è definita come la somma di Trend (μ_t), Ciclo (ψ_t), Stagionalità (γ_t) e White Noise (ε_t):

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

Alcune di queste componenti possono essere non presenti ed altre aggiunte. inoltre, possono essere viste come una versione stocastica della funzione deterministica del tempo.

2.1 Trend

Il *Trend* è responsabile della variazione della media del processo nel lungo periodo. Questa componente solitamente viene usata come ***L***ocal ***L***inear ***T***rend (**LLT**). Si considera la seguente funzione lineare:

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

con μ_0 intercetta e β_0 la pendenza e la si scriva nella sua forma incrementale:

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

Aggiungendo White Noise η_n è possibile ottenere Random Walk con Drift. In questo caso μ_t è interpretabile com un trend lineare con una intercetta Random Walk, mentre la pendenza rimane costante.

Inoltre, è possibile ottenere una pendenza che varia nel tempo facendo evolvere 'intercetta come un Random Walk:

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

Queste equazioni definiscono un Local Linear Trend interpretabili come un trend lineare nella quale sia intercetta che pendenza evolvono come Random Walk.

Il Local Linear Trend presenta diversi casi d'interesse aggiustando il valore delle varianze:

- *Deterministic Linear Trend* se $\sigma_\eta^2 = \sigma_\xi^2 = 0$;
- *Random Walk con Drift* β_0 se $\sigma_\xi^2 = 0$ (pendenza costante);
- *Random Walk* se $\sigma_\xi^2 = \beta_0 = 0$ (pendenza = 0);
- *Integrated Random Walk* se $\sigma_\eta^2 = 0$, con un trend molto lineare.

Obs.: il Local Linear Trend può essere visto com un modello ARIMA se $\sigma_\xi^2 > 0$, $\mu_t \sim I(2)$ con trend non stazionario, ma la sua seconda differenza stazionaria.

2.2 Forma State Space

Una serie storica può essere scritta in forma *State Space*, un sistema di equazioni nella quale una o più serie storiche osservabili sono linearmente correlate con un set di variabili non osservabili. Essa è definita nel seguente modo:

$$\begin{cases} Y_t & = d_t + Z_t\alpha_t + \varepsilon_t \quad \text{Equazione di Osservazione} \\ \alpha_{t+1} & = c_t + T_t\alpha_t + R_t\eta_t \quad \text{Equazione di Stato} \end{cases}$$

dove $\varepsilon_t \sim WN(0, H_t)$, $\eta_t \sim WN(0, Q_t)$ sono delle variabili casuali incorrelate tra di loro con valore atteso nullo e matrice di varianza covarianza rispettivamente H_t e Q_t.

Il processo di inizializzazione è un processo nella quale il vettore di stato è definito come $a_{1|0} = \mathbb{E}(\alpha_1)$ e $P_{1|0} = \mathbb{E}(\alpha_1 - a_{1|0})(\alpha_1 - a_{1|0})^t$.

Esempio. Si consideri la trasformazione in formato State Space del Local Linear Trend con White Noise . Il Local Linear Trend presenta il seguente sistema di equazioni:

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

con rumore $y_t = \mu_t + \varepsilon_t$. La forma State Space del Local Linear Trend diventa:

$$\begin{cases} \alpha_{t+1} = T\alpha_t + R\eta_t = \begin{bmatrix} \mu_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix} & \text{Equazione di Stato} \\ Y_t = Z\alpha_t + \varepsilon_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \varepsilon_t & \text{Equazione di Osservazione} \end{cases}$$

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

Il processo di inizializzazione diventa:

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

Oss. L'Integrated Random Walk è un Local Linear Trend con la varianza del primo disturbo pari a 0.

2.2.1 Filtro di Kalman

Si suppone che tutti i parametri della forma State Space siano noti. Le uniche componenti non note sono quelle non osservabili, specificate come variabili casuali, e l'inferenza si basa sulla previsione statistica attraverso:

- Forecsting, forma State Space α_t basata su Y_s ($s < t$);
- Filtering, forma State Space α_t basata su Y_t;
- Smoothing, forma State Space α_t basata su Y_s ($s > t$).

In particolare, è possibile costruire uno stimatore ottimo lineare migliore di ogni predittore assumendo che ε_t, $v_t \sim WN$ e l'initial state α_1 sono congiuntamente Gaussiane.

Si consideri la seguente notazione:

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

Il *Filtro di Kalman* è un algoritmo utilizzato per calcolare la coppia $\{a_{t|t-1}, P_{t|t-1}\}$ a partire da $\{a_{t-1|t-1}, P_{t-1|t-1}\}$ e $\{a_{t|t}, P_{t|t}\}$ a partire da $\{a_{t|t-1}, P_{t|t-1}\}$. Perciò proietta le stime di y_t partendo basate sulle osservazioni precedenti.

Inoltre fornisce le sequenze di innovazioni con la rispettiva matrice di varianza covarianza, utilizzata per stimare la verosimiglianza Gaussiana del modello nella forma State Space.

- 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:

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

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

La previsione lineare del vettore di stato α_t basato su $y_s = \{Y_1, \dots, Y_s\}$ è chiamato *Smoothing*. Lo Smoother aggiustato a livello di intervalli fornisce una previsione per il vettore di stato basato sulla intera serie storica:

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

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

Il Disturbance Smoothing calcola la previsione delle sequenze White Noise basate sulla intera serie storica ed è utile per identificare possibili outliers:

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

Se le osservazioni sono Normalmente distribuite, il Filtro di Kalman permette di costruire la funzione di verosimiglianza. Utilizzando la definizione di densità condizionata, è possibile fattorizzare la densità congiunta dei dati:

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

Sotto queste assunzioni, la distribuzione condizionata è anch'essa Normale ed il Filtro di Kalman ha media e varianza come segue:

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

La verosimiglianza è la densità congiunta del percorso del campione dei parametri non noti θ:

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

Generalmente è più utile utilizzate la log-verosimiglianza:

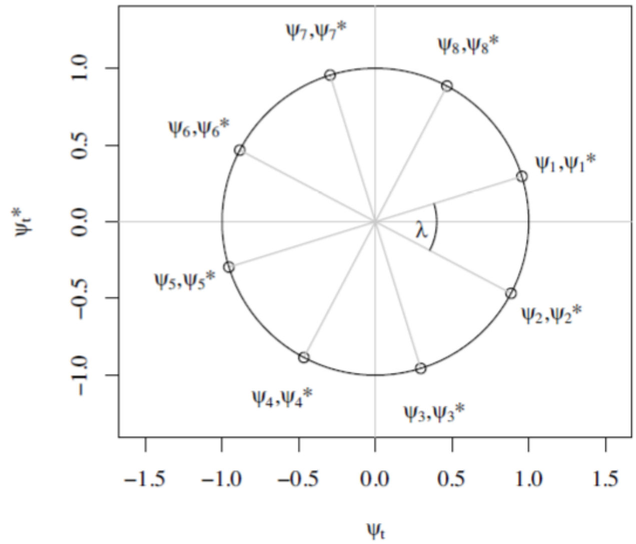
$$\begin{aligned} l(\theta) &= \log f(y_1, \dots, y_n) = \sum_{t=1}^{n} \log f(y_t|y_1, \dots, y_{t-1}) = \\ &= \sum_{t=1}^{n} -1/2(\log|F_t(\theta)| + \\ &= (y_t - \hat{y}_{t|t-1}(\theta)^t F_t(\theta)^{-1}(y_t - \hat{y}_{t|t-1}(\theta)^t \\ &\implies \hat{\theta}_n = \max l(\theta) \end{aligned}$$

2.3 Ciclo Stocastico

Un ciclo si verifica quando i dati mostrano aumenti e diminuzioni che non sono di una frequenza fissa. La funzione deterministica per la costruzione di un *Ciclo* di frequenza λ è la sinusoide $R\cos(\lambda t + \phi)$ dove R è chiamata Amplitudine e ϕ chiamata Fase. Un modo per generare in termini geometrici la sinusoide è individuare una circonferenza di raggio R il punto della fase. Il ciclo ψ_t è stazionario:

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

dove $\rho \in [0, 1]$ è definito damping factor, $\lambda \in [0, \pi]$ la frequenza del ciclo e k_t, k^*_t sono sequenze indipendenti tra loro White Noise con varianza σ^2_k.



Esiste un caso unico nella quale il processo ha soluzioni stazionarie e casuali $\mathbb{E}[\psi_t] = 0$, $\mathbb{E}[\psi_{t+h}\psi^t_t] = \frac{\sigma^2_k}{1-\rho^2} I^2$ e $\psi_t \sim ARMA(2, 1)$ con radici complesse nel polinomio AR.

Se $\rho = 1$ e $\mathbb{E}[\psi_t] = 0$, il ciclo non è stazionario, mentre se $\rho < 1$ il ciclo è stazionario.

2.4 Stagionalità

La componente *Stagionale* si manifesta nel momento in cui la serie storica presenta dei fattori stagionali come il giorno di una settimana. Essa riguarda sempre una frequenza fissa e nota. La componente stagionale presente nei modelli UCM è generalmente modellata tramite un forma dummy stocastica in termini trigonometrici.

Si considera la seguente serie storica con stagionalità s:

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

La *Forma Trigonometrica* è data da $\gamma_t = \sum_{j=1}^{s/2} \gamma^{(j)}_t$, con $\gamma^{(j)}_t$ ciclo non stazionario stocastico:

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

Gli argomenti della sinusoide ono chiamate frequenze stagionali. Se s è pari e $j = s/2$, la seconda equazione del precedente sistema può essere omesso in quanto $\sin(\pi) = 0$ e la prima equazione può essere ridotta a

$\gamma_{t+1}^{(s/2)} = -\gamma_t^{(s/2)} - w_t^{(s/2)}$ che non dipende dai valori della seconda equazione.

La forma Dummy Stocastica è un altro modo di modellazione della componente stagionale definendo s variabili che evolvono come RW. Sia γ_t l'effetto stagionale al tempo t:

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

dove $w_t \sim WN(0, \sigma_w^2)$. Il modo più veloce per far evolvere la funzione in modo stocastico è aggiungere dei random shock di media 0.