Dipartimento di Matematica
Corso di Laurea Magistrale in Matematica Applicata

Rielaborazione degli appunti del corso di

Analisi di Sequenza Dati

Simmaco Di Lillo
dsimmaco@gmail.com

a.a. 21-22

Introduzione

Queste note contengono i miei appunti personali presi durante il corso di "Analisi di sequenza dati" del Prof. C. Cammarota e dunque non sono dispense ufficiali del corso.
Nelle note potrebbero essere presenti typo o errori, per qualsiasi segnalazione scrivetemi un'e-mail a dsimmaco@gmail.com, la versione aggiornata verrà caricata sul mio sito

Indice

1 Lezione del 7 Marzo

Definizione 1.1. Sia F la funzione di ripartizione di una variabile aleatoria continua X. Il **quantile** è la funzione inversa di F

Definizione 1.2. Date due distribuzioni empiriche il comando `qqplot`

- crea un griglia di valori nell'intervallo 0.1
- Trova i quantili corrispondenti per entrambe le distribuzioni
- Disegna nel piano cartesiano le coppie di punti

Osservazione 1. Più i punti sono allineati con la bisettrice del primo e terzo quadrante più le due distribuzioni coincidono

Osservazione 2. Data una distribuzione empirica il comando `qqnorm` calcola media e deviazione standard della distribuzione empirica. Confronta la distribuzione empirica con una gaussiana di stessa media e deviazione standard.

1.1 Stima dei parametri di una distribuzione normale

Sia $X \sim \mathcal{N}(m, \sigma^2)$ e sia $X_1, \dots, X_n$ un campione casuale estratto da X

1.1.1 Stima di m data σ

Supponiamo $\sigma = \sigma_0$ noto. Lo stimatore

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

è corretto di varianza $\dfrac{\sigma^2}{n}$.

Poichè $\sqrt{n}\dfrac{\overline{X}_n - m}{\sigma_0}$ è una normale standard si ha

$$1 - \alpha = \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < \sqrt{n}\frac{\overline{X}_n - m}{\sigma_0} < z_{1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(X - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < m < X + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right)$$

dunque

$$\left(X - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, X + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right)$$

è un intervallo di confidenza per m noto σ_0 di livello $1 - \alpha$

Osservazione 3. Se $\alpha = 0,005$ allora $z_{1-\frac{\alpha}{2}} = 1,96 \approx 2$

1.1.2 Stima di σ nota m

Supponiamo $m = m_0$ nota. Consideriamo lo stimatore

$$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - m_0)^2$$

Ora

$$\frac{X_i - m}{\sigma} \sim \mathcal{N}(0,1) \quad \Rightarrow \quad X_i - m = \sigma Z_i \text{ con } Z_1, \dots, Z_n \sim \mathcal{N}(0,1) \text{ i.i.d}$$

dunque

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma^2 Z_i^2 = \frac{\sigma^2}{n}\chi^2(n)$$

Poichè $\mathbb{E}\left[\chi^2(n)\right] = \sum_{i=1}^{n} \mathbb{E}\left[Z_i^2\right] = n$ S_n^2 è uno stimatore corretto di σ^2.

$$1-\alpha = \mathbb{P}\left(\chi^2_{n,\frac{\alpha}{2}} < \frac{n}{\sigma^2}S_n^2 < \chi^2_{n,1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(\frac{n}{\chi^2_{n,1-\frac{\alpha}{2}}}S_n^2 < \sigma^2 < \frac{n}{\chi^2_{n,\frac{\alpha}{2}}}S_n^2\right)$$

dunque

$$\left(\frac{n}{\chi^2_{n,1-\frac{\alpha}{2}}}S_n^2, \frac{n}{\chi^2_{n,\frac{\alpha}{2}}}S_n^2\right)$$

è un intervallo di confidenza per σ^2 noto m di livello $1-\alpha$

1.1.3 Stima di σ^2 con m non noto

Consideriamo lo stimatore

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

Chiamiamo $V_i = X_i - m$ si ha $\mathbb{E}\left[V_i\right] = 0$ e

$$X_i - \overline{X}_n = V_i + m - \frac{1}{n}\sum_{i=1}^{n}(V_i + m) = V_i + \frac{1}{n}\sum_{i=1}^{n} V_i$$

Da cui

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(V_i - \frac{1}{n}\sum_{j=1}^{n} V_j\right)^2 = \frac{1}{n-1}\sum_{i}\left(V_i^2 - \frac{2}{n}V_i\sum_{j} V_j + \frac{1}{n^2}\sum_{j,k} V_j V_k\right)$$

Ora sfruttando l'indipendenza di V_i, V_j per $i \neq j$ e poichè le V_i hanno media nulla, otteniamo

$$\mathbb{E}\left[S_n^2\right] = \frac{1}{n-1}\sum_{i}\left(\sigma^2 - \frac{2}{n}\sigma^2 + \frac{1}{n^2}n\sigma^2\right) = \sigma^2$$

Dal teorema di Cochran si dimostra il seguente

Corollario 1.1.

$$S_n^2 \sim \frac{\sigma^2}{n-1}\chi^2(n-1)$$

da cui con calcoli analoghi al caso di m nota otteniamo

$$\left(\frac{n-1}{\chi^2_{n-1,1-\frac{\alpha}{2}}}S_n^2, \frac{n-1}{\chi^2_{n-1,\frac{\alpha}{2}}}S_n^2\right)$$

è un intervallo di confidenza per σ^2 di livello $1-\alpha$

1.1.4 Stima di m

Consideriamo lo stimatore

$$T = \frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} \sim \sqrt{n}\frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n-1}\chi^2(n-1)}} = \frac{\overline{X}_n - m}{\frac{\sigma}{\sqrt{n}}}\frac{1}{\sqrt{\frac{\chi^2(n-1)}{n-1}}} \sim t(n-1)$$

Dunque

$$\left(\overline{X}_n - \frac{t_{n-1,1-\frac{\alpha}{2}}}{\sqrt{n}}S_n^2, \overline{X}_n + \frac{t_{n-1,1-\frac{\alpha}{2}}}{\sqrt{n}}S_n^2\right)$$

è un intervallo di confidenza per m di livello $1-\alpha$

1.2 Test di significatività

Consideriamo un'ipotesi base H_0 che per motivi storici prende il nome di ipotesi nulla. Tale ipotesi, di solito, è di tipo semplificatico e conservativo (deve mantenere lo "status quo": ci vuole una forte evidenza per rigettare la situazione attuale).
Fissato un valore soglia α, una classe di test si compone delle seguenti fasi:

- Si costruisce un indicatore detto statistica del test
- Si stabilisce la distribuzione della statistica in H_0
- Si confronta il valore empirico osservato T con la distribuzione. Se il valore appartiene ad una certa regione (che dipende da α) si rifiuta l'ipotesi nulla.

Definizione 1.3. Il p-value è il più piccolo valore di α per cui i dati campionari consentono di rifiutare l'ipotesi nulla.

Esempio 1.2 (Test di comparazione tra le medie di due campioni)**.** *Siano $X_i \sim \mathcal{N}(m_X, \sigma_X^2)$ n v.a. i.i.d e $Y_i \sim \mathcal{N}(m_y, \sigma_y^2)$ m v.a. i.i.d.*
Cerchiamo un test per l'ipotesi nulla $H_0 : m_X = m_Y$.
Definiamo la varianza raggruppata come

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

allora la statistica test

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

è distribuita come $t(n+m-2)$. Il test è dunque rifiuto se $|T| > t_{n+m-2,1-\frac{\alpha}{2}}$

2 Lezione del 14 Marzo

Definizione 2.1. Un **dataframe** è una tabella di dati dove

- le righe sono le unità di campionamento
- le colonne sono le grandezze misurate (osservate)

2.1 Regressione lineare

Lo scopo del modello lineare è quello di stimare una variabile dipendente detta target usando delle variabili indipendenti detti predittori o covariate.
Fissiamo la notazione

- y è la variabile target (colonna di osservazioni)
- X matrice avente per colonne le variabili predittori
- ω è il rumore
- β sono i coefficienti detti parametri

Dunque supponendo $\omega \sim \mathcal{N}\left(0, \sigma^2 I\right)$ con σ incognito cerchiamo β che verifichi

$$y = X\beta + \omega \tag{1}$$

2.1.1 Metodo dei minimi quadrati

Cerco $\hat{\beta}$ che minimizza $||y - X\beta||^2$ dunque deve valere

$$\begin{aligned}
0 &= \frac{\partial}{\partial \beta_s} \sum_{i=1}^{n} \left(y_i - (X\beta)_i\right)^2 = -\sum_{i=1}^{n} 2\left(y_i - (X\beta)_i\right) \frac{\partial}{\partial \beta_s}(X\beta)_i = \\
&= -2\sum_{i=1}^{n} \left(y_i - (X\beta)_i\right) \frac{\partial}{\partial \beta_s}\left(\sum_{k=1}^{q} X_{ik}\beta_k\right) = -2\sum_{i=1}^{n}\left(y_i - (X\beta)_i\right) X_{is} = \\
&= -2\sum_{i=1}^{n} X^T_{si} y_i - X^T_{si}(X\beta)_i = -2\left(\left(X^T y\right)_s - \left(X^T X\beta\right)_s\right) \; \forall s = 1, \dots q
\end{aligned}$$

dunque deve accadere $X^T y = \left(X^T X\right)\beta$ e assumendo $X^T X$ invertibile otteniamo

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T y$$

Osservazione 4. La richiesta sull'invertibilità della matrice segue dall'invertibilità di X. Se non lo fosse potrei non considerare un sottoinsieme di colonne linearmente indipendenti.

Proposizione 2.1. *Il vettore q-dimensionale*

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2\left(X^T X\right)^{-1}\right)$$

dunque è uno stimatore non distorto.

Dimostrazione. Usando l'equazione (1) e la definizione di $\hat{\beta}$ otteniamo

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T \left(X\beta + \omega\right) = \left(X^T X\right)^{-1} X^T X\beta + \left(X^T X\right) X^T \omega = \beta + \left(X^T X\right)^{-1} X^T \omega$$

dunque $\hat{\beta}$ è una normale di media

$$\mathbb{E}\left[\hat{\beta}\right] = \beta + \left(X^T X\right)^{-1} X^T \mathbb{E}\left[\omega\right] = \beta$$

Ricordando che $Cov(AZ) = ACov(Z)A^T$ otteniamo

$$\begin{aligned} Cov(\hat{\beta}) &= Cov\left(\beta + \left(X^T X\right)^{-1} X^T \omega\right) = Cov\left(\left(X^T X\right)^{-1} X^T \omega\right) = \\ &= \left(X^T X\right)^{-1} X^T \sigma^2 I \left(\left(X^T X\right)^{-1} X^T\right)^T = \sigma^2 \left(X^T X\right)^{-1} X^T X \left(X^T X\right)^{-1} = \sigma^2 \left(X^T X\right)^{-1} \end{aligned}$$

□

A questo punto, supponendo σ noto abbiamo

$$\frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{\left(X^T X\right)^{-1}_{ij}}} \sim \mathcal{N}(0, 1)$$

e dunque è possibile costruire un intervallo di confidenza per β_j di livello $1 - \alpha$

Definizione 2.2. Il vettore dei **valori predetti** (*fitted values*) è

$$\hat{y} = X\hat{\beta} = X \left(X^T X\right)^{-1} X^T y = Hy$$

dove

$$H = X \left(X^T X\right)^{-1} X^T$$

prende il nome di **matrice cappello**.

Il vettore dei **residui** è

$$e = y - \hat{y}$$

Proposizione 2.2 (Proprietà di H)**.**

(i) $H^2 = H$

(ii) H *è simmetrica*

(iii) $Tr(H) = q$

Dimostrazione. (i) e (ii) sono semplici manipolazioni algebriche. Per (iii) basta usare il fatto che $Tr(AB) = Tr(BA)$ quando ha senso scambiare l'ordine di moltiplicazione.

Lemma 2.3. *Sia A una matrice deterministica e u un vettore aleatorio. $u^t A u$ è una forma quadratica aleatoria e vale*

$$\mathbb{E}\left[u^T A u\right] = \mathbb{E}\left[u^T\right] A \mathbb{E}\left[u\right] + tr(AC)$$

dove C è la matrice di covarianza di u

Dimostrazione.

$$\mathbb{E}\left[u^T A u\right] = \mathbb{E}\left[\sum_{i,j} A_{ij} u_i u_j\right] = \sum_{i,j} A_{ij} \mathbb{E}\left[u_i u_j\right] = \sum_{i,j} A_{ij}\left(\mathbb{E}\left[u_i\right]\mathbb{E}\left[u_j\right] + Cov(u_i, u_j)\right) =$$
$$= \mathbb{E}\left[u^T\right] A \mathbb{E}\left[u\right] + \sum_{i,j} A_{ij} C_{ij} = \mathbb{E}\left[u^T\right] A \mathbb{E}\left[u\right] + Tr(AC)$$

dove l'ultima uguaglianza deriva dalla simmetria della matrice C

$$\sum_{ij} A_{ij} C_{ij} = \sum_{ij} A_{ij} C_{ji} = \sum_i \sum_j A_{ij} C_{ji} = \sum_i (AC)_i = Tr(AC)$$

□

Proposizione 2.4. *I marginali e_i sono normali con media 0 e varianza $\sigma^2(1 - H_{ii})$ mentre*

$$\mathbb{E}\left[||e||^2\right] = (n - q)\sigma^2$$

Dimostrazione.

$$\mathbb{E}\left[y - \hat{y}\right] = \mathbb{E}\left[X\beta + \omega - X\hat{\beta}\right] = X\beta - X\beta = 0$$

Per quanto riguarda la covariannza notiamo che $e = y - Hy$ e dunque

$$Cov(e) = Cov((I - H)y) = (I - H)Cov(y)(I - H)^T = \sigma^2(I - H)^2 = \sigma^2(I - H)$$

dove abbiamo usato l'idenpotenza di H.
Per la seconda parte:

$$E\left[||e||^2\right] = \mathbb{E}\left[||(I - H)\,y||^2\right] = \mathbb{E}\left[((I - H)y)^T\,(I - H)\,y\right] = \mathbb{E}\left[y^T(I - H)^2 y\right]$$

ora sfruttando l'idenpotenza di H otteniamo

$$\begin{aligned} E\left[||e||^2\right] &= \mathbb{E}\left[y^T\,(I - H)\,y\right] = \mathbb{E}\left[y^T\right](I - H)\mathbb{E}\left[y\right] + Tr\left((I - H)\,Cov(y)\right) \\ &= (X\beta)^T(I - H)(X\beta) + \sigma^2 Tr(I - H) = \beta^T X^T X\beta - \beta^T X^T H X\beta + \sigma^2(n - q) = \\ &= \beta^T X^T X\beta - \beta^T X^T\left(X\left(X^T X\right)^{-1} X^T\right)X\beta + \sigma^2(n - q) = \\ &= \beta^T X^T X\beta - \beta^T\left(X^T X\right)\left(X^T X\right)^{-1} X^T X\beta + \sigma^2(n - q) = \sigma^2(n - q) \end{aligned}$$

3 Lezione del 21 Marzo

I residui sono definiti come

$$e = y - \hat{y}$$

dove $\hat{y} = Hy$ sono i “fitted value" e vale

$$e \sim \mathcal{N}\left(0, \sigma^2(I - H)\right) \quad \Rightarrow \quad e_i \sim \mathcal{N}\left(0, \sigma^2(1 - H_{ii})\right)$$

dunque

$$\frac{e_i}{\sigma\sqrt{1 - H_{ii}}} \sim Z$$

L'analisi dei residui consiste nel verificare che i residui, opportunamente normalizzati, sono distribuiti come una Z.
Se σ viene stimato da

$$\hat{\sigma} = \frac{1}{n-q}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n-q}\,||e||^2$$

allora definiamo gli “studentized residual" come

$$\frac{e_i}{\hat{\sigma}\sqrt{1 - H_{ii}}} \sim t(n-q)$$

dove

- n è il numero di osservazioni e
- q è il numero di predittori

Osservazione 5. Se $n - q > 7$ possiamo approssimare la distribuzione di Student come una normale

Osservazione 6. I residui devono essere contenuti tranne eccezionali casi tra -3 e 3 ed inoltre devono essere invarianti per cambio di indice.

Andiamo a fare inferenza sui coefficienti di regressione.
Abbiamo provato che

$$Cov\left(\hat{\beta}\right) = \sigma^2\left(X^tX\right)^{-1}$$

dunque

$$st.dev\left(\hat{\beta}_j\right) = \sigma\sqrt{(X^tX)^{-1}_{ii}}$$

e un suo stimatore è

$$st.\hat{d}ev\left(\hat{\beta}_j\right) - \hat{\sigma}\sqrt{(X^tX)^{-1}_{ii}}$$

Il test che voglio condurre ha come ipotesi nulla

$$H_0 : \ \beta_j = 0$$

ovvero β_j è un finto predittore.
Dunque un intervallo di confidenza per β_j di livello $1 - \alpha$ è

$$\hat{\beta}_j \pm st.\hat{d}ev\left(\hat{\beta}_j\right) t_{1-\frac{\alpha}{2}, n-q}$$

Osservazione 7. Se

$$\hat{\beta}_j > 2st.\hat{d}ev\left(\hat{\beta}_j\right)$$

allora rifiuto l'ipotesi nulla: il valore stimato è significativo. Sotto tale ipotesi 0 non appartiene all'intervallo di confidenza.

Proposizione 3.1. *I residui sono ortogonali ai valori fittati*

Dimostrazione.

$$e^t\hat{y} = (y-\hat{y})^t\,\hat{y} = [(I-H)y]^t\,Hy = y^t(I-H)Hy = y^t\left(H-H^2\right)y = 0$$

dove l'ultima uguaglianza segue da $H^2 = H$

Corollario 3.2.

$$||y||^2 = ||\hat{y}||^2 + ||e||^2$$

3.1 Predizione

Supponiamo di avare una nuova osservazione x_0 definiamo

- la predizione della risposta singola come

  $$y_0 = x_0^t\hat{\beta} + \omega_0$$

  dove ω_0 è un nuove termine di rumore

- la predizione della risposta media come

  $$\hat{y}_0 = x_0^t\hat{\beta}$$

Notiamo che vale

$$E\left[y_0\right] = x_0^t\beta$$

$$Var(y_0) = Var\left(x_0^t\hat{\beta}\right) + Var(\omega_0) = x_0^T Cov\left(\hat{\beta}\right)x_0 + \sigma^2 = x_0^t\sigma^2\left(X^tX\right)^{-1}x_0 + \sigma^2 = \sigma^2\left(1 + x_0^t\left(X^tX\right)^{-1}x_0\right)$$

Esempio 3.3 (Caso di un solo predittore)**.** *In questo caso la matrice X è una singola colonna fatta di* 1.

$$\hat{\beta} = \left(X^tX\right)^{-1}X^Ty = \frac{1}{n}\sum_{y_i} := \overline{y}$$

Definizione 3.1. Definiamo la "sum of square error" di ordine come

$$SSE_q = \sum_{i=1}^{n}\left(y_i - \hat{y}_i^{(q)}\right)^2$$

dove $\hat{y}^{(q)}$ sono i valori fittati con il modello a q predittori.
Assumendo $SSE_q < SSE_1$ definiamo il coefficiente di determinazione come

$$R_q^2 = \frac{SSE_1 - SSE_q}{SSE_1}$$

Osservazione 8. Poichè assumiamo che la prima colonna di X sia formata da tutti 1

$$SSE_1 = \sum_{i=1}^{n}\left(y_i - \hat{y}_i^{(1)}\right)^2 = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$$

Osservazione 9. Il coefficiente R^2 ci dice quanto migliora l'errore se usiamo q predittori invece che 1

Per fare inferenza su R^2 osserviamo che

$$F = \frac{\frac{R_q^2}{q-1}}{\frac{1-R_q^2}{n-q}} \sim F(q-1, n-q)$$

dove F è la distribuzione di Fisher

$$F(n, m) \sim \frac{\frac{\chi^2(n)}{n}}{\frac{\chi^2(m)}{m}}$$

con le χ^2 indipendenti.

3.2 Applicazioni alle serie temporali

- Modello trend + rumore:
$$X_t = f(t) + \omega_t$$
dove
  - ω_t i.i.d $(0, \sigma^2)$ e
  - $f(t)$ è un trend deterministico

  In questo caso X_t è una sequenza di v.a indipendenti con valore atteso $f(t)$ e varianza σ^2.

- Modello trend+stagionalità + rumore
$$X_t = f(t) + S_t + \omega_t$$
dove S_t è una sequenza deterministica tale che
$$\sum_{t=0}^{p-1} S_t = 0 \qquad S_{t+p} = S_t$$
Un modo per scegliere S_t è utilizzare la regressione armonica.
Data una pulsazione ω supporre
$$S_t = \beta_c \cos(\omega t) + \beta_s \sin(wt)$$
e usare una regressione lineare con colonne $\cos(\omega t)$ e $\sin(\omega t)$.
Se T è il periodo, possiamo scegliere $\omega = \frac{2\pi}{T}$

4 Lezione del 28 Marzo

4.1 LEZIONE MANCANTE

5 Lezione del 4 Aprile

Definizione 5.1. Dato uno stimatore T di t definiamo le seguenti quantità

- La varianza
$$Var(T) = \mathbb{E}\left[(T - \mathbb{E}[T])^2\right]$$
- La distorsione
$$Bias(T) = \mathbb{E}[T] - t$$
- Il rischio
$$Risk(T) = \mathbb{E}\left[(T - t)^2\right]$$

Lemma 5.1. *Dato uno stimatore T di t vale*

$$Risk(T) = Var(T) + Bias(T)^2$$

Dimostrazione.

$$\begin{aligned} Risk(T) &= \mathbb{E}\left[(T-t)^2\right] = \mathbb{E}\left[(T - \mathbb{E}[T] + \mathbb{E}[T] - t)^2\right] = \\ &= \mathbb{E}\left[(T - \mathbb{E}[T])^2 + (\mathbb{E}[T] - t)^2 - 2(T - \mathbb{E}[T])(\mathbb{E}[T] - t)\right] = \\ &= Var(T) + Bias(T)^2 + 0 \end{aligned}$$

□

5.1 Media mobile

Supponiamo di avere un processo stocastico della forma trend+ rumore $X_t = f(t) + \omega_t$. Consideriamo lo stimatore

$$\hat{f}(t) = \sum_{j=-L}^{L} a_j X_{t-j} \text{ dove } \sum_{j=-L}^{L} a_j = 1$$

allora

$$\mathbb{E}\left[\hat{f}(t)\right] = \sum_{j=-L}^{L} a_j f(t-j)$$

$$Var\left(\hat{f}(t)\right) = \sum_{j=-L}^{L} a_j^2 Var\left(X_{t-j}\right) = \sigma^2 \sum_{j=-L}^{L} a_j^2$$

Esempio 5.2 (Boxcar)**.** *Se prendiamo $a_j = \dfrac{1}{2L+1}$ per $j = -L, \ldots, L$ allora*

$$Var(\hat{f}(t)) = \frac{\sigma^2}{2L+1}$$

dunque all'aumentare di L, la varianza dello stimatore diminuisce. Ma, se il trend varia, usando un L troppo grande potrei perdere il trend: aumenta la distorsione.

Esempio 5.3 (Boxcar a trend lineare)**.** *Supponiamo che $f(t) = \beta_0 + \beta_1 t$ allora lo stimatore ottenuto è corretto $\forall L$*

Dimostrazione.

$$\mathbb{E}\left[\hat{f}(t)\right] = \sum_{j=-L}^{L} a_j f(t-j) = \sum_{j=-L}^{L} \frac{1}{2L+1}\left(\beta_0 + \beta_1(t-j)\right) = \beta_0 + \beta_1 t = f(t)$$

dunque se ho un trend lineare è conveniente prendere un L grande per diminuire la varianza

Esempio 5.4 (Filtro per trend quadratico)**.** *Supponiamo* $f(t) = \beta_0 + \beta_1 t + \beta_2 t^2$ *allora prendendo*

$$a_0 = \frac{1}{2}, \quad a_1 = a_{-1} = \frac{1}{3}, \quad a_2 = a_{-2} = -\frac{1}{12}$$

si ottiene uno stimatore corretto

Dimostrazione.

$$\begin{aligned}\sum_{j=-L}^{L} a_j f(t-j) &= \sum_{j=-L}^{L} a_j \left(f(t) - jf'(t) + \frac{j^2}{2} f''(t)\right) = \\ &= \sum_{j=-L}^{L} a_j f(t) - \sum_{j=-L}^{L} a_j j\left(\beta_1 - 2\beta_2\right)t + \sum_{j=-L}^{L} a_j j^2 \beta_2 = \\ &= f(t) \sum_{j=-L}^{L} a_j - (\beta_1 + 2\beta_2)t \sum_{j=-L}^{L} a_j j + \beta_2 \sum_{j=-L}^{L} j^2 a_j\end{aligned}$$

dunque deve valere

$$\begin{cases} \sum_{j=-L}^{L} a_j = 1 \\ \sum_{j=-L}^{L} j a_j = 0 \\ \sum_{j=-L}^{L} j^2 a_j = 0 \end{cases}$$

prendendo $L = 2$ e a_i come nell'ipotesi il sistema viene risolto □

Osservazione 10. Con metodi analoghi si calcolano i coefficienti per trend polinomiali di grado qualsiasi. Per i polinomi di grado 12 si usa un filtro a 15 valori noto come filtro di Spencer

Esempio 5.5 (Boxcar a trend quadratico)**.**

$$Var(\hat{f}(t)) = \frac{\sigma^2}{2L+1}$$

$$\begin{aligned}Bias(\hat{f}(t)) &= \mathbb{E}\left[\hat{f}(t)\right] - f(t) = \mathbb{E}\left[\sum_j (a_j X_{t-j} + a_j w_{t-j})\right] - f(t) = \\ &= \sum_j a_j f(t-j) - f(t) = \frac{1}{2L+1}\sum_j (f(t-j) - f(t)) = \\ &= \sum_j \left(f'(t)(-j) + \frac{j^2}{2} f''(t)\right) = \frac{1}{2L+1}\left(-f'(t)\sum_j j + \frac{f''(t)}{2}\sum_j j^2\right) \approx \\ &\approx \frac{\beta_2}{2L+1}\int_{-L}^{L} x^2 \, \mathrm{d}x = \frac{\beta_2}{2L+1} \cdot \frac{2L^3}{3} \approx \frac{\beta_2}{3} L^2\end{aligned}$$

da cui

$$Risk\left(\hat{f}(t)\right) \approx \frac{\sigma^2}{2L} + \frac{\beta_2^2}{9}L^4$$

volendo minimizzarlo impongo

$$\frac{\partial}{\partial L}Risk = 0 \quad \Leftrightarrow \quad -\frac{\sigma^2}{2L^2} + \frac{4\beta_2^2}{9}L^3 = 0 \quad \Leftrightarrow \quad L = \left(\frac{9\sigma^2}{16\beta_2^2}\right)^{\frac{1}{5}}$$

5.2 Stabilizzazione della varianza

Supponiamo di avere un modello trend+rumore a varianza non costante. Ad esempio consideriamo il caso

$$X_t = f(t) + \sigma(t)z_t \qquad z_t = i.i.d(0,1)$$

Cerco una funzione $U : \mathbb{R} \to \mathbb{R}$ tale che $U(X_t)$ abbia varianza costante

$$U\left(X_t\right) = U\left(f(t) + \sigma(t)z_t\right) \approx U(f(t)) + U'(f(t))\sigma(t)z_t$$

$$Var(U(X_t)) = \left(U'(f(t))\sigma(t)\right)^2$$

Esempio 5.6. *Se $\sigma(t) = kf(t)$ allora per quanto visto in precedenza*

$$Var(U(X_t)) = c^2 \quad \Leftrightarrow \quad U'^2(x)k^2x^2 = c^2 \quad \Leftrightarrow \quad U'(x)kx = c \quad \Leftrightarrow \quad U(x) = \frac{c}{k}\log x \ \textit{se}\ x > 0$$

Osservazione 11. Se $\sigma(t) = \sqrt{f(t)}$ allora $U(x) = \sqrt{x}$ stabilizza la varianza

6 Lezione del 11 Aprile

Definizione 6.1. Un processo stocastico (X_t) si dice stazionario se per ogni $k \in \mathbb{N}$ la distribuzione congiunta di k variabili è invariante per traslazione

Osservazione 12. Ricordiamo che se un processo X_t è stazionario allora

$$\mathbb{E}[X_t] = m \quad Var(X_t)\sigma^2 \quad \forall t$$

ed inoltre

$$Cov(X_t, X_s) = \gamma(|t-s|)$$

Definizione 6.2. Se valgono solamente le proprietà dell'osservazione diremo che il processo è stazionario in senso debole

Osservazione 13. Se X_t è gaussiano allora le due definizioni coincidono

Osservazione 14. Il modello trend+ rumore non è stazionario. $\mathbb{E}[X_t] = f(t)$ dipende da t in generale

Definizione 6.3. Una passeggiata casuale o moto Browniano è un processo X_t dove

$$X_t = \sum_{j=1}^{t} \omega_j \text{ dove } \omega_j = i.i.d(0, \sigma^2)$$

Osservazione 15. Il moto browniano non è stazionario

$$\mathbb{E}[X_t] = \sum_{j=1}^{t} \mathbb{E}[\omega_j] = 0$$

Ma

$$Var(X_t) = \sum_{j=1}^{t} Var(\omega_j) = \sigma^2 t$$

dunque la varianza cresce al crescere di t. Al crescere di t, le varie realizzazioni “tagliano" una sezione maggiore dell'area orizzontale.

Proposizione 6.1.

$$Cov(X_t, X_s) = \sigma^2 \min\{t, s\}$$

Dimostrazione.

$$Cov(X_t, X_s) = \mathbb{E}\left[\sum_{j=1}^{t} \omega_j \sum_{k=1}^{s} \omega_k\right] = \sum_{j=1}^{t}\sum_{k=1}^{s} \mathbb{E}[\omega_j w_k] = \sum_{j=1}^{\min\{t,s\}} \mathbb{E}[w_j^2] = \sigma^2 \min\{t, s\}$$

Osservazione 16. Se X_t è una passeggiata casuale allora $\Delta X_t = X_t - X_{t-1} = \omega_t$

Definizione 6.4. Una passeggiata casuale con deriva (drift) è un processo stocastico X_t della forma

$$X_t = \sum_{j=1}^{t} \omega_j \text{ dove } \omega_j = i.i.d(\mu, \sigma^2)$$

dove $\mu \neq 0$

Proposizione 6.2 (Differenza di rumore)**.** *Sia* $\omega_t = i.i.d(0, \sigma^2)$ *e definiamo il processo*

$$X_t = \Delta\omega_t = \omega_t - \omega_{t-1}$$

allora ottengo un proccesso stazionario con funzione di autocorrelazione

$$\rho(h) = \begin{cases} 1 & se\ h = 0 \\ -\frac{1}{2} & se\ |h| = 1 \\ 0 & altrimenti \end{cases}$$

Dimostrazione.

$$\mathbb{E}[X_t] = \mathbb{E}[\omega_t] - \mathbb{E}[\omega_{t-1} = 0$$
$$Var(X_t) = Var(\omega_t) + Var(\omega_{t-1}) = 2\sigma^2$$

Sia $h > 0$ allora

$$Cov(X_t, X_{t+h}) = \mathbb{E}\left[\omega_t\omega_{t+h} - \omega_{t-1}\omega_{t_h} - \omega_t\omega_{t+h-1} + \omega_{t-1}\omega_{t+h-1}\right]$$

Ora se $h > 1$ si ha $Cov(X_t, X_{t+h}) = 0$ mentre se $h = 1$ otteniamo $-\sigma^2$.

Definizione 6.5. Data una processo stocastico X_t definiamo

$$r_t = \frac{X_t}{X_{t-1}}$$

e i logaritmic returns come

$$\log r_t = \log X_t - \log X_{t-1} = \Delta \log X_t$$

Esempio 6.3. *Siano* $A, B = i.i.d(0, \sigma^2)$ *e consideriamo il processo*

$$X_t = A\cos\omega t + B\sin\omega t$$

Allora il processo è debolmente stazionario con funzione di autocovarianza $\rho(h) = \cos\omega h$

Dimostrazione. Mostriamo che è debolmente stazionario

$$\mathbb{E}[X_t] = 0$$

$$Var(X_t) = \sigma^2 cos^2\omega t + \sin^2\omega t = \sigma^2$$
$$\gamma(h) = \mathbb{E}[X_t X_{t+h}] = \mathbb{E}\left[(\ldots)(\ldots)\right] = \mathbb{E}\left[A^2\cos\omega t\cos\omega(t+h) + B^2\sin\omega t\sin\omega(t+h)\right]$$

Ricordando che $\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$ otteniamo

$$\gamma(h) = \sigma^2\cos(\omega t - \omega(t+h)) = \sigma^2\cos\omega h$$

da cui

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \cos\omega h$$

6.1 Predizione del valore di una serie stazionaria con l'autocovarianza

Assumiamo che valga

$$X_{t+h}^{p} = \alpha X_t$$

allora

$$\mathbb{E}\left[\left(X_{t+h}^{p} - X_{t+h}\right)^2\right] = \mathbb{E}\left[\alpha^2 X_t^2 - 2\alpha X_t X_{t+h} + X_{t+h}^2\right] = \alpha^2\sigma^2 - 2\alpha\gamma(h) + \sigma^2$$

Volendo minimizzare l'errore impongo

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\alpha^2\sigma^2 - 2\alpha\gamma(h) + \sigma^2\right) = 0 \quad \Leftrightarrow \quad \alpha = \rho(h)$$

per tale α l'errore di predizione vale

$$\begin{aligned}\mathbb{E}\left[\left(X_{t+h}^{p} - X_{t+h}\right)^2\right] &= \rho(h)^2\sigma^2 - 2\rho(h)\gamma(h) + \sigma^2 = \\ &= \rho(h)^2\sigma^2 - 2\rho(h)^2\sigma^2 + \sigma^2 = \sigma^2\left(1 - \rho(h)^2\right)\left(\gamma(h) - \rho(h)\sigma^2\right)\end{aligned}$$

6.2 Processi ARIMA

6.2.1 Processi Autoregressive

Definizione 6.6. Un processo X_t si autoregressivo di ordine p, abbreviato $AR(p)$ se è stazionario in senso debole e soddisfa per opportuni $\phi_1, \dots, \phi_p$ l'equazione:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \omega_p$$

dove X_t è stazionario e $\omega_t = i.i.d(0, \sigma^2)$ e indipendente da X_j con $j < t$ e X_t.
Il termine stocastico ω_t prende il nome di innovazione (per sottolineare l'indipendenza da tutte le variabili X_j precedenti)

Osservazione 17. Possiamo scrivere il processo in una forma matriciale

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_t \end{pmatrix} = \begin{pmatrix} X_0 & X_{-1} & \dots & X-p \\ X_1 & X_0 & \dots & X_{1-p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ X_{t-1} & X_{t-2} & \dots & X_{t-p} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} + \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \vdots \\ \omega_t \end{pmatrix}$$

La forma matriciale ci dice anche che i coefficienti ϕ_i possono essere stimati con la regressione lineare.

Esempio 6.4 ($AR(1)$)**.**

$$X_t = \phi_1 X_{t-1} + \omega_t$$

Se X_t è stazionario e $|\phi_1| \leq 1$ vale

$$\rho(h) = \phi_1^{|h|}$$

dunque in valore assoluto decresce esponenzialmente

Dimostrazione.

$$Var(X_t) = \phi_1^2 Var(X_{t-1}) + \sigma^2$$

dunque poichè il processo è stazionario

$$\gamma(0) = \phi_1^2 \gamma(0) + \sigma^2 \quad \Leftrightarrow \quad \gamma(0) = \frac{\sigma^2}{1 - \phi_1^2}$$

Sia $h > 0$ allora

$$\gamma(h) = \mathbb{E}[X_t X_{t+h}] = \mathbb{E}[X_t\,(\phi_1 X_{t+h-1} + \omega_{t+h})] = \phi_1 \mathbb{E}[X_t X_{t+h-1}] + \mathbb{E}[X_t w_{t+h}] = \phi_1 \gamma(h-1)$$

Dunque abbiamo provato che

$$\gamma_h = \phi_1^h \gamma(0)$$

7 Lezione del 2 Maggio

Proposizione 7.1 (Equazione di Yule-Walker per l'autocovarianza)**.** *Sia X_t un processo $AR(p)$ allora*

$$\gamma(h) = \phi_1\gamma(h-1) + \cdots + \gamma_p\gamma(h-p)$$

Dimostrazione. I processi AR hanno media nulla (??) dunque

$$\gamma(h) = \mathbb{E}[X_t X_{t+h} = \mathbb{E}\left[X_t\left(\phi_1 X_{t+h-1} + \cdots + \phi_p X_{t+h-p}\right)\right] = \phi_1\gamma(h-1) + \cdots + \gamma_p\gamma(h-p)$$

Esempio 7.2 ($AR(2)$)**.** *Trovare un'espressione per $\gamma(h)$ nel caso di un processo $AR(2)$*

Dimostrazione. Dalla proposizione precedente

$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2)$$

Risolvendo l'equazione per ricorrenza (cerco una soluzione della forma $\gamma(h) = z^h$) ottengo i seguenti casi

- Se $\phi_1^2 + 4\phi_2 > 0$ allora il polinomio caratteristico ha due radici distinte dunque la soluzione ha la forma
$$\gamma(h) = c_1 z_1^h + c_2 z_2^h \qquad c_1, c_2 \in \mathbb{R}$$
dove
$$z_{1,2} = \frac{2\phi_2}{-\phi_1 \pm \sqrt{\phi_1^2 + 2\phi_2}}$$

- Se $\phi_1^2 + 4\phi_2 < 0$ il polinomio caratteristico ha due radici complesse coniugate $z = re^{\pm i\theta}$ dunque
$$\gamma_h = r^h\left(c_1\cos\theta h + c_2\sin\theta h\right) \qquad c_1, c_2 \in \mathbb{R}$$

Per calcolare i coefficienti c_1, c_2 usiamo le condizioni iniziali $\gamma(0)$ e $\gamma(1)$ infatti

$$\gamma(1) = \phi_1\gamma(0) + \phi_2\gamma(-1) \quad \Rightarrow \quad \gamma(1) = \frac{\phi_1}{1-\phi_2}\gamma(0)$$

Ma

$$\gamma(0) = Var(X_t) = \mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[\left(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \omega_t\right)^2\right]$$

e svolgendo il quadrato otteniamo

$$\gamma(0) = \phi_1^2\gamma(0) + \phi_2^2\gamma(0) + \sigma^2 + 2\phi_1\phi_2\gamma(1)$$

dunque ricordando l'espressione per $\gamma(1)$ otteniamo

$$\gamma(0) = \frac{1-\phi_2}{1+\phi_2}\cdot\frac{\sigma^2}{(1-\phi_2^2)-\phi_1^2}$$

7.1 Processi a media mobile

Definizione 7.1. Un processo so dice a media mobile di ordine q, abbreviato $MA(q)$ se è della forma

$$X_t = \sum_{j=0}^{q} \theta_j \omega_{t-j} \quad \text{con } \theta_0 = 1$$

dove $\omega_t = i.i.d(0, \sigma^2)$

Esempio 7.3 ($MA(1)$)**.**

$$X_t = \omega_t + \theta_1 \omega_{t-1}$$

Allora vale

$$\rho(h) = \begin{cases} 1 & se\ h = 0 \\ \dfrac{\theta_1}{1+\theta_1^2} & se\ |h| = 1 \\ 0 & altrimenti \end{cases}$$

Dimostrazione.

$$Var(X_t) = Var(\omega_t + \theta_1 \omega_{t-1}) = \mathbb{E}\left[\omega_t^2 + \omega_{t-1}^2 + 2\theta_1 \omega_t \omega_{t-1}\right] = (1+\theta_1^2)\sigma^2$$

Sia $h > 0$ allora

$$\begin{aligned}\gamma(h) = Cov\left(X_t, X_{t+h}\right) = \mathbb{E}\left[X_t X_{t+h}\right] = \mathbb{E}\left[\left(\omega_t + \theta_1 \omega_{t-1}\right)\left(\omega_{t+h} + \theta_1 \omega_{t+h-1}\right)\right] = \\ = \mathbb{E}\left[\omega_t \omega_{t+h} + \theta_1 \omega_t \omega_{t+h-1} + \theta_1 \omega_{t-1}\omega_{t+h} + \theta_1^2 \omega_{t-1}\omega_{t+h-1}\right]\end{aligned}$$

Se $i > j$ allora $\mathbb{E}[\omega_i \omega_j] = \mathbb{E}[\omega_i]\mathbb{E}[\omega_j] = 0$ dunque

$$\gamma(1) = \theta_1 \sigma^2$$

$$\gamma(h) = 0 \text{ se } h > 1$$

dunque dividendo per $Var(X_t)$ otteniamo la tesi □

Osservazione 18. Poichè

$$\frac{\theta_1}{1+\theta_1^2} = \frac{\frac{1}{\theta_1}}{1 + \left(\frac{1}{\theta_1}\right)^2}$$

per eliminare l'ambiguità si sceglie $|\theta_1| \leq 1$

Osservazione 19. Se $\theta_1 = 1$ riotteniamo la differenza di rumore

Esempio 7.4 ($MA(q)$)**.** *Sia*

$$X_t = \sum_{j=0}^{q} \theta_j \omega_{t-j}$$

allora

$$\rho(h) = \begin{cases} 1 & se\ h = 0 \\ \dfrac{\sum \theta_j \theta_{j+h}}{\sum \theta_i^2} & se\ h = 1 \\ 0 & altrimenti \end{cases}$$

Dimostrazione.

$$Var(X_t) = \mathbb{E}\left[\left(\sum_{j=0}^{q}\theta_j\omega_{t-j}\right)^2\right] = \mathbb{E}\left[\left(\sum_j\theta_j\omega_{t-j}\right)\left(\sum_k\theta_k\omega_{t-k}\right)\right] =$$
$$= \sum_j\sum_k\theta_j\theta_k\mathbb{E}\left[\omega_{t-j}\omega_{t-k}\right] = \sigma^2\sum_j\theta_j^2$$

Sia $h > 0$ allora

$$\gamma(h) = \mathbb{E}\left[X_tX_{t+h}\right] = \mathbb{E}\left[\left(\sum_j\theta_j\omega_{t-j}\right)\left(\sum_k\theta_k\omega_{t+h-k}\right)\right] =$$
$$= \sum_j\sum_k\theta_i\theta_k\mathbb{E}\left[\omega_{t-j}\omega_{t+h-k}\right] = \sum_{j=0}^{q-h}\theta_j\theta_{j+h}\sigma^2$$

7.2 Processi ARMA

Definizione 7.2. Un processo si dice $ARMA(p,q)$ se è della forma

$$X_t = \phi_1X_{t-1} + \cdots + \phi_pX_{t-p} + \omega_t + \theta_1\omega_{t-1} + \theta_q\omega_{t-q}$$

Osservazione 20. Un processo $ARMA(0,0)$ è semplicemente rumore.

Definizione 7.3. Detto $\Delta X_t = X_t - X_{t-1}$ l'operatore differenza.
Diremo che un processo X_t è $ARIMA(p,d,q)$ se Δ^dX_t è $ARMA(p,q)$.

Osservazione 21. I processi $ARIMA$ sono, in generale, non stazionari

Osservazione 22. Un processo $ARIMA(0,1,0)$ è una passeggiata casuale

Definizione 7.4. Definiamo l'operatore di Back-Shift come

$$BX_t = X_{t-1}$$

Osservazione 23. Con l'introduzione di B, possiamo scrivere

$$\Delta = I - B$$

Definizione 7.5. Un processo si dice causale se può essere scritto nella forma

$$X_t = \sum_{j=0}^{\infty}\psi_j\omega_{t-j}$$

dove (ψ_j) è una famiglia sommabile e $\omega_t = i.i.d(0,\sigma^2)$

Proposizione 7.5. *$AR(1)$ è causale*

Dimostrazione. Euristica

$$X_t = \phi_1X_{t-1} + \omega_t \quad\Rightarrow\quad X_t - \phi_1X_{t-1} = \omega_t \quad\Rightarrow\quad (I - \phi_1B)X_t = \omega_t \quad\Rightarrow\quad X_t = \frac{1}{1-\phi_1B}\omega_t$$

Ora ricordando che

$$\frac{1}{1-\phi z} = \sum_{j=0}^{\infty}(\phi_1z)^j \text{ se } |\phi_1| \le 1$$

Dimostrazione. Formale. Proviamo che

$$Z_t := \sum_{j=0}^{\infty} \phi_1^j B^j w_t = \sum_{j=0}^{\infty} \phi_1^j \omega_{t-j}$$

verifica $Z_t = \omega_t + \phi_1 Z_{t-1}$ il che dimostra che $Z_t = X_t$

$$Z_t = \sum_{j=0}^{\infty} \phi_1^j \omega_{t-j} = \phi_1^0 \omega_{t-0} + \sum_{j=1}^{\infty} \phi_1^j \omega_{t-j} = \omega_t + \sum_{j=0}^{\infty} \phi_1^{j+1} \omega_{(t-1)-j} = \omega_t + \phi_1 Z_{t-1}$$

□

8 Lezione del 9 Maggio

Osservazione 24. Se X_t ammette una rappresentazione causale:

$$X_t = \sum_{j=0}^{\infty} \psi_j \omega_{t-j}$$

dove $\omega_{t-j} = i.i.d(0, \sigma^2)$ allora

$$\mathbb{E}[X_t] = 0$$
$$Var(X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$$
$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$$

dunque il processo è debolmente stazionario

Dimostrazione.
$$\mathbb{E}[X_t] = \sum_j \psi_j \mathbb{E}[\omega_{t-j}] = 0$$
$$Var(X_t) = \mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[\sum_{j=0}^{\infty} \psi_j \omega_{t-j} \sum_{k=0}^{\infty} \psi_k \omega_{t-k}\right] \sum_j \sum_k \psi_j \psi_k \mathbb{E}\left[\omega_{t-j}\omega_{t-k}\right] = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$$

Sia $h > 0$ allora

$$\gamma(h) = \mathbb{E}\left[X_t X_{t+h}\right] = \mathbb{E}\left[\sum_j \psi_j \omega_{t-j} \sum_k \psi_k \omega_{t+h-k}\right] = \sum_j \sum_k \psi_j \psi_k \mathbb{E}\left[\omega_{t-j}\omega_{t+h-k}\right]$$

Ora il valore atteso è diverso da 0 solo se $k = j + h$ da cui la tesi □

Proposizione 8.1. *Un processo $AR(p)$ è causale.*

Dimostrazione. Euristica. Sia X_t un processo $AR(p)$ dunque esistono dei coefficienti ϕ_i tali che

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \omega_t$$

detto $\phi(B) = I - \phi_1 B - \cdots - \phi_p B^p$ l'equazione precedente si scrive come

$$\phi(B) X_t = \omega_t \quad \Rightarrow \quad X_t = \frac{1}{\phi(B)} \omega_t$$

Ora detto $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z$, siano z_j le sue radici con $|z_j| > 1$ dunque:

$$\phi(z) = \prod_{j=1}^{p} \left(1 - \frac{z}{z_j}\right)$$

da cui applicandolo alla serie otteniamo

$$\begin{aligned} X_t &= \prod_{j=1}^{p} \frac{I}{I - \frac{B}{z_j}} \omega_t = \sum_{k_1=0}^{\infty} \cdots \sum_{k_p=1}^{\infty} \frac{B^{k_1}}{z_1^{k_1}} \cdots \frac{B^{k_p}}{z_p^{k_p}} \omega_t = \sum_{k_1=0}^{\infty} \cdots \sum_{k_p=1}^{\infty} z_1^{-k_1} \cdots z_p^{-k_p} \omega_{t-\sum k_i} \\ &= \sum_{j=0}^{\infty} \left(\sum_{\substack{k_1, \ldots k_p \\ \sum k_i = j}} z_1^{-k_1} \cdots z_p^{-k_p} \right) w_{t-j} \end{aligned} \tag{2}$$

Dimostrazione. Formalmente basta dimostrare che la serie definita da

$$\sum_{j=0}^{\infty} \psi_j \omega_{t-j} \text{ con } \psi_j = \sum_{\substack{k_1,\dots k_p \\ \sum k_i = j}} z_1^{-k_1} \cdots z_p^{-k_p}$$

verifica l'equazione del modello $AR(p)$

Proposizione 8.2. *I processi $ARMA(p,q)$ sono causali*

Dimostrazione. Si pone $\Theta(B) = I + \theta_1 B + \cdots + \theta_q B^q$ dunque osserviamo che

$$\phi(B)X_t = \Theta(B)\omega_t \quad \Rightarrow \quad X_t = \frac{\Theta(B)}{\phi(B)}\omega_t$$

Si procede, dunque, come nella dimostrazione precedente sostitudendo nell'equazione (2) ω_t con $\Theta(B)\omega_t$

Esempio 8.3. *Troviamo una rappresentazione causale di $ARMA(1,1)$*

Dimostrazione. Gli operatori definiti in precedenza valgono

$$\phi(B) = I - \phi_1 B \qquad \Theta(B) = I + \theta_1 B$$

dunque

$$\begin{aligned} X_t =& \frac{I + \theta_1 B}{I - \phi_1 B}\omega_t = (I + \theta_1 B)\left(\sum_{j=0}^{\infty} \phi_1^j B^j w_t\right) = \sum_{j=0}^{\infty} \phi_1^j B^j w_j + \sum_{j=0}^{\infty} \theta_1 \phi_1^j B^{j+1}\omega_t = \\ =& \omega_t + \sum_{j=1}^{\infty} \phi_1^j \omega_{t-j} + \sum_{j=1}^{\infty} \theta_1 \phi_1^{j-1}\omega_{t-j} = \omega_t + \sum_{j=1}^{\infty} \phi_1^{j-1}\left(\theta_1 + \phi_1\right)\omega_{t-j} \end{aligned} \tag{3}$$

dunque ponendo

$$\psi_j = \begin{cases} 1 & \text{se } j = 0 \\ \phi_1^{j-1}\left(\theta_1 + \phi_1\right) & \text{se } \geq 1 \end{cases}$$

otteniamo la rappresentazione causale

$$X_t = \sum_{j=0}^{\infty} \psi_j \omega_{t-j}$$

Esercizio 8.4. *Calcolare la varianza di un processo $ARMA(1,1)$ e la funzione di autocorrelazione*

Dimostrazione. Per farlo possiamo usare la rappresentazione causale

$$Var(X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \sigma^2 + \sigma^2 \sum_{j=1}^{\infty} \phi_1^{2(j-1)}\left(\phi_1 + \theta_1\right)^2 = \sigma^2 + \frac{\left(\sigma(\phi_1 + \theta_1)\right)^2}{1 - \phi_1^2}$$

Usando la formula dell'osservazione 24 otteniamo

$$\rho(h) = \frac{(\phi_1 + \theta_1)(1 + \theta_1 \psi_1)}{1 + 2\theta_1 \psi_1 + \theta_1^2}\phi_1^{h-1} \qquad \text{per } h \geq 1$$

8.1 Stima dei parametri

Per stimare i parametri del modello $AR(p)$ possiamo utilizzare il modello lineare.
La cosa è più difficile per i modelli $MA(q)$

Definizione 8.1. Un processo X_t si dice avere una rappresentazione invertibile se $\exists(\pi_j)_j$ sommabile tale che

$$\sum_{j=0}^{\infty} \pi_j X_{t-j} = \omega_t$$

Proposizione 8.5. *$MA(1)$ ha una rappresentazione invertibile*

Dimostrazione. Euristica

$$X_t = \omega_t + \theta_1 \omega_{t-1} = (I + \theta_1 B)\omega_t \quad \Rightarrow \quad \omega_t = \frac{1}{I + \theta_1 B} X_t = \sum_{j=0}^{\infty} (-\theta_1)^j X_{t-j}$$

Per i modelli MA la stima dei parametri si effettua con il metodo dei momenti

Esempio 8.6 ($MA(1)$)**.** *Dq quanto osservato vale*

$$\begin{array}{ll} \textit{Media:} & \mu = 0 \\ \textit{Varianza:} & \gamma(0) = (1+\theta_1^2)\sigma^2 \\ \textit{Autcorrelazione:} & \rho(h) = \begin{cases} 1 & \textit{se } h = 0 \\ \frac{\theta_1}{1+\theta_1^2} & \textit{se } |h| = 1 \\ 0 & \textit{altrimenti} \end{cases} \end{array}$$

Poniamo ora

$$\hat{\gamma}(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} \left(X_t - \overline{X}\right)\left(X_{t+h} - \overline{X}\right)$$

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Le equazioni sono dunque

$$\begin{cases} \hat{\sigma} = \hat{\gamma}(0) \\ \hat{\theta}_1 = \frac{\hat{\rho}(1)}{1+\hat{\rho}(1)} \end{cases}$$

Per l'analisi dei residui si utilizza il test di Ljung–Box che ha per ipotesi nulla $\rho(h) = 0$ per $h = 1, \dots, H$
Tale test si basa sulla statistica

$$Q = n(n+2) \sum_{h=1}^{H} \frac{\hat{\rho}^2(h)}{n-h}$$

tale test deriva dal Box-Test che usava la statistica

$$Q = n \sum_{h=1}^{H} \hat{\rho}(h)$$

distribuita asintoticamente come una χ^2_H infatti $\hat{\rho}(h) \sim A\mathcal{N}\left(0, \frac{1}{n}\right)$

9 Lezione del 16 Maggio

9.1 Analisi spettrale

Definizione 9.1. La trasformata discreta di Fourier (DFT) è la funzione

$$\mathcal{F} : \mathbb{C}^n \to \mathbb{C}^n \qquad \mathcal{F}(X) = d$$

definita nel seguente modo

$$d_j = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} e^{-i2\pi \frac{j}{N} t} X_t \text{ per } j = 0, \dots, N-1$$

e chiameremo

- $f_j = \dfrac{j}{N}$ frequenze armonicche
- $f_1 = \dfrac{1}{N}$ frequenza fondamentale
- se N pari $f_{\frac{N}{2}} = \dfrac{1}{2}$ frequenza di taglio

Osservazione 25. La trasformata discreta è indotta dalla matrice di entrate

$$\mathcal{F}_{jt} = \frac{1}{\sqrt{N}} e^{-i2\pi \frac{j}{N} t}$$

che è simmetrica e ortonormale

Dimostrazione.

$$\langle \mathcal{F}_{j\cdot}, \mathcal{F}_{k\cdot} \rangle = \sum_{t=0}^{N-1} \mathcal{F}_{jt} \overline{\mathcal{F}}_{k,t} = \sum_t \frac{1}{\sqrt{N}} e^{-i2\pi \frac{j}{N} t} \frac{1}{\sqrt{N}} e^{i2\pi \frac{k}{N} t} = \frac{1}{N} \sum_t e^{-i2\pi \frac{(j-k)}{N} t}$$

Ora se $j = k$ poichè $e^0 = 1$ vale $\langle \mathcal{F}_{j\cdot}, \mathcal{F}_{j\cdot} \rangle = 1$.
Sia $j \neq k$ allora ponendo

$$z = e^{-i2\pi \frac{j-k}{N}}$$

e notando $z \neq 1$ infatti j, k non superano N, otteniamo

$$\langle \mathcal{F}_{j\cdot}, \mathcal{F}_{k\cdot} \rangle = \sum_{t=0}^{N-1} z^t = \frac{1}{N} \frac{1 - z^N}{1 - z} = 0$$

Osservazione 26. Dall'osservazionne precedente notiamo che la trasformata è invertibile. Vale

$$X_t = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{i2\pi \frac{j}{N} t} d_j$$

Osservazione 27. Essendo la DFT ortonormale conserva la norma:

$$||d||^2 = ||X||^2$$

Osservazione 28. Se X ha media nulla allora $||X||^2$ è una devianza. La DFT è una "decomposizione della varianza"

Definizione 9.2. Dato $X \in \mathbb{C}^n$ e posto $d = \mathcal{F}X$, il periodogramma di X è

$$I_j = |d_j|^2$$

Osservazione 29.

$$I_0 = |d_0|^2 = \left| \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} e^{i2\pi \frac{0}{N} t} X_t \right|^2 = \frac{1}{N} \left(N\overline{X} \right)^2 = N\overline{X}^2$$

Proposizione 9.1. *Se $X \in \mathbb{R}^N$ allora $I_{N-j} = I_j$: il periodograma è simmetrico rispetto a $\frac{N}{2}$*

Dimostrazione.

$$I_{N-J} = d_{N-j}\overline{d}_{N-j} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} e^{-i2\pi \frac{N-j}{N} t} X_t \frac{1}{\sqrt{N}} \sum_{s=0}^{N-1} e^{i2\pi \frac{N-j}{N} s} \overline{X_s}$$

Scomponendo

$$e^{-i2\pi \frac{N-j}{N} t} = e^{-i2\pi t} e^{i2\pi \frac{j}{N} t} = e^{i2\pi \frac{j}{N} t}$$

ottengo

$$I_{N-J} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} e^{-i2\pi \frac{j}{N} t} X_t \frac{1}{\sqrt{N}} \sum_{s=0}^{N-1} e^{i2\pi \frac{j}{N} s} X_s = \overline{d}_j d_j = I_j$$

Esempio 9.2. *Sia*

$$X_t = Ae^{i2\pi \frac{k}{N} t}$$

allora

$$d_j = \begin{cases} A\sqrt{N} & \textit{se } j = k \\ 0 & \textit{altrimenti} \end{cases}$$

infatti

$$d_j = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} e^{-2\pi \frac{j}{N} t} Ae^{i2\pi \frac{k}{N} t} = \frac{A}{\sqrt{N}} \sum_{t=0}^{N-1} e^{-i2\pi \frac{j-k}{N} t}$$

Osservazione 30 (Relazione tra periodogramma e autocovarianza).

$$I_j = \sum_{h=-(N-1)}^{N-1} e^{-i2\pi \frac{j}{N} h} \hat{\gamma}(h)$$

infatti

$$I_J = d_j\overline{d}_j = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} e^{-i2\pi \frac{j}{N} t} X_t \frac{1}{\sqrt{N}} \sum_{s=0}^{N-1} e^{i2\pi \frac{j}{N} s} X_s = \sum_{h=-(N-1)}^{N-1} e^{-i2\pi \frac{j}{N} h} \frac{1}{N} \sum_{\substack{s,t=0 \\ t-s=h}}^{N-1} X_t X_s$$

Osservazione 31 (Periodogramma del rumore gaussiano). Sia $X_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d. Allora

$$d_j = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \cos\left(2\pi \frac{j}{N} t\right) X_t - \frac{i}{\sqrt{N}} \sum_{t=0}^{N-1} \sin\left(2\pi \frac{j}{N} t\right) X_t = c_j - is_j$$

dove c_j si chiama "cosine transformate" e s_j "sine transformate".
Dunque

$$I_j = |d_j|^2 = c_j^2 + s_j^2$$

Ora $E[c_j] = 0$ mentre

$$Var(c_j) = \frac{1}{N} \sum \cos^2 \left(2\pi \frac{j}{N} t\right) \sigma^2 = \frac{\sigma^2}{N} \sum \cos^2 \left(2\pi \frac{j}{N} t\right) \approx \frac{\sigma^2}{2}$$

analogamente per s_j dunque possiamo approssimare

$$\frac{\sqrt{2}}{\sigma} c_j \approx Z_1 \sim \mathcal{N}(0,1) \qquad \frac{\sqrt{2}}{\sigma} s_j \approx Z_2 \sim \mathcal{N}(0,1)$$

Inoltre c_j, s_j si possono approssimare come v.a scorrelate infatti

$$\begin{aligned} \mathbb{E}[c_j s_j] &= \mathbb{E}\left[\frac{1}{N} \sum_t \cos\left(2\pi \frac{j}{N} t\right) \omega_t \sum_s \sin\left(2\pi \frac{j}{N} s\right) \omega_s\right] = \\ &= \frac{1}{N} \sum_{t,s} \cos\left(2\pi \frac{j}{N} t\right) \sin\left(2\pi \frac{j}{N} s\right) \mathbb{E}[\omega_t \omega_s] = \frac{1}{N} \sum_t \cos\left(2\pi \frac{j}{N} t\right) \sin\left(2\pi \frac{j}{N} t\right) \sigma^2 \approx 0 \end{aligned}$$

Da cui $I_j \approx \chi^2(2) = \mathcal{E}\left(\frac{1}{2}\right)$

10 Lezione del 23 Maggio

Definizione 10.1. Sia X_t un processo stocastico e supponiamo che la sua funzione di autocovarianza $\gamma(h)$ sia sommabile. La densità spettrale è definita come

$$I(\omega) = \sum_{h \in \mathbb{Z}} e^{i\omega h} \gamma(h)$$

Osservazione 32. Nel limite per $N \to +\infty$ e sostituendo a $2\pi f_i$ la quantià $\omega \in [0, 2\pi]$ otteniamo dalla definizione di trasformata discreta di Fourie, la formulazione della densità spettrale.

Osservazione 33. Dalla teoria sulle serie di Fourier vale

$$\gamma(h) = \int_0^{2\pi} e^{i\omega h} I(\omega) \, \mathrm{d}\omega$$

Esempio 10.1. *Se $X_t = \mathcal{N}(0, \sigma^2)$ i.i.d. allora*

$$I(\omega) = \sigma^2 \quad \forall \omega \in [0, 2\pi]$$

Esempio 10.2 (AR(1))**.** *La densità spettrale di un processo $AR(1)$ vale*

$$I(\omega) = \frac{\sigma^2}{1 + \phi_1^2 - 2\pi_1 \cos\omega}$$

Dimostrazione. Sappiamo che

$$\gamma(h) = \frac{\phi_1^{|h|}}{1 - \phi_1^2} \sigma^2$$

da cui

$$\begin{aligned} I(\omega) =& \sum_{h \in \mathbb{Z}} e^{-i\omega h} \frac{\phi_1^{|h|}}{1 - \phi_1^2} \sigma^2 = \frac{\sigma^2}{1 - \phi_1^2} \left(\sum_{h \geq 0} e^{-i\omega h} \phi_1^h + \sum_{h > 0} e^{i\omega h} \phi_1^h \right) = \\ =& \frac{\sigma^2}{1 - \phi_1^2} \left(\frac{1}{1 - \phi_1 e^{i\omega}} + \frac{\phi_1 e^{i\omega}}{1 - \phi_1 e^{i\omega}} \right) = \frac{\sigma^2}{1 - \phi_1^2} \frac{1 - \phi_1^2}{|1 - e^{i\omega}\phi_1|^2} = \\ =& \frac{\sigma^2}{1 - \phi_1^2} \cdot \frac{1 - \phi_1^2}{(1 - \phi_1 \cos\omega)^2 + (\phi_1 \sin\omega)^2} = \frac{\sigma^2}{1 + \phi_1^2 - 2\phi_1 \cos\omega} \end{aligned}$$

Osservazione 34. Se $\phi_1 > 0$ allora $I(0) > I(\pi)$ dunque le frequenza basse intervengono maggiormente.
Se $\phi_1 < 0$ invece $I(0) < I(\pi)$ e dunque hanno maggior peso le frequenze alte

Definizione 10.2. Un filtro è una sequenza $(a_k)_{k>0}$. Data una sequenza X_t, la sequenza filtrata è

$$Y_t = \sum_{k=0}^{t} a_k X_{t-k}$$

Proposizione 10.3. *Dato un filtro (a_k) vale*

$$D^Y(\omega) = d^X(\omega) A(\omega)$$

dove

$$A(\omega) = \sum_{k=0}^{\infty} e^{-i\omega k} a_k$$

e prende il nome di funzione di trasferimento del filtro.

Dimostrazione. Partiamo dal caso discreto

$$d_j^Y = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{2i\pi \frac{j}{N} t} Y_t = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} e^{2i\pi \frac{j}{N} t} \sum_{k=0}^{t} a_k X_{t-k} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} e^{-2i\pi \frac{j}{N}(t-k)} e^{-2i\pi \frac{j}{N}} \sum_{k=0}^{\infty} a_k X_{t-k} =$$
$$= \sum_{k=0}^{N-1} \frac{1}{\sqrt{N}} \sum_{t=k}^{N-1} e^{-2i\pi \frac{j}{N}(t-k)} X_{t-k} e^{-2i\pi \frac{j}{N} k} a_k = \sum_{k=0}^{N-1} e^{-2i\pi \frac{j}{N} k} a_k \frac{1}{\sqrt{N}} \sum_{t=0}^{N-k-1} e^{-2i\pi \frac{j}{N} t} X_t$$

Nel limite $N \to +\infty$ e $2\pi \frac{j}{N} \to \omega$ otteniamo la tesi □

Corollario 10.4. *Sia Y la filtrata di X tramite il filtro (a_k) allora vale*

$$I_Y(\omega) = I_X(\omega) \left| A(\omega) \right|^2$$

Esempio 10.5 (Densità spettrale di processi $ARMA(p,q)$)**.** *Sotto opportune ipotesi vale*

$$I_X(\omega) = \frac{\left| \Theta\left(e^{-i\omega}\right) \right|^2}{\left| \Phi\left(e^{-i\omega}\right) \right|^2} \sigma^2$$

dove Φ, Θ sono gli usuali polinomi.

Dimostrazione. Sia

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=0}^{q} \theta_j \omega_{t-j} \text{ con } w_j = i.i.d(0, \sigma^2)$$

allora usando gli usuali polinomi formali otteniamo

$$X = \frac{\Theta(B)}{\Phi(B)} \omega \quad \Rightarrow \quad X_t = \sum_{j=0}^{\infty} \psi_j \omega_{t-j}$$

da cui

$$I_X(\omega) = \left| \Psi(\omega) \right|^2 \sigma^2 \text{ dove } \Psi(\omega) = \sum_{j=0}^{\infty} e^{-ij\omega} \psi_j$$

infatti la densità del rumore vale costantemente σ^2.
Se vale

$$\frac{\Theta(z)}{\Phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j$$

allora

$$\Psi(\omega) = \frac{\Theta\left(e^{-i\omega}\right)}{\Phi\left(e^{-i\omega}\right)} \quad \Rightarrow \quad I_X(\omega) = \frac{\left| \Theta\left(e^{-i\omega}\right) \right|^2}{\left| \Phi\left(e^{-i\omega}\right) \right|^2} \sigma^2$$