

Metodi Statistici per l'Analisi Socio-Economica

7. Processi stocastici



L'approccio moderno alle serie storiche

Spesso non si riesce a modellare in modo soddisfacente il processo generatore dei dati come $Y_t = f(t) + u_t$. In questo caso è meglio omettere la parte deterministica e concentrarsi sulla parte stocastica:

- si sposta l'attenzione su u_t
- u_t non viene più assunto *white noise*
- ... bensì si ipotizza che $Cov(u_t, u_s) \neq 0$ per qualche t, s

Non si cerca più di scomporre la serie in componenti. Sfruttando la correlazione insita nel processo, si cerca di individuare un modello probabilistico che descriva l'evoluzione del fenomeno.



I processi stocastici

Definizione formale: consideriamo uno spazio di probabilità (Ω, F, P) dove

- Ω è lo spazio degli eventi elementari
- F una σ -algebra su Ω
- P una misura di probabilità

dato uno spazio parametrico τ , si definisce *processo stocastico* una funzione finita e a valori reali di $\omega \in \Omega$ e $t \in \tau$ tale che, per ogni t , $Y_t(\omega)$ è una funzione misurabile di ω .

D'ora in poi,

- t indicherà il *tempo*
- τ sarà discreto

I processi stocastici: traiettorie

Un processo stocastico è pertanto funzione di due variabili: un indice temporale t che dà un ordine alla famiglia di v.c., e ω un evento che specifica su Ω quale risultato si verifica per ogni t .

Inoltre:

- per ogni fissato $\omega = \omega_0$, $Y_t(\omega_0)$ è una funzione di t detta *realizzazione* o *traiettoria* del processo stocastico.
- per ogni $t = t_0$ fissato, $Y_{t_0}(\omega)$ è una funzione misurabile di $\omega \in \Omega$ e quindi una variabile casuale
- per $t = t_0$ e $\omega = \omega_0$, $Y_{t_0}(\omega_0)$ è un numero reale

Lo stesso processo stocastico può pertanto generare traiettorie (*realizzazioni finite*) diverse.

I processi stocastici: descrizione

Semplificando, si può dunque dire che un processo stocastico è *una collezione di variabili casuali indicizzate dal tempo*

$$\{Y_t, t = 1, 2, \dots, n\}$$

Per la sua descrizione completa, servirebbe la distribuzione congiunta di $(Y_{t_1}, \dots, Y_{t_n})$ per ogni insieme di istanti t_1, \dots, t_n ed ogni n . Di solito ci si limita a descrivere il processo per mezzo dei momenti di ordine 1 e 2 delle Y_t , che al variare di t sono le seguenti funzioni a valori reali:

- funzione *media*: $\mu_t = E[Y_t]$
- funzione *varianza*: $\sigma_t^2 = \text{Var}[Y_t] = E[(Y_t - \mu_t)^2]$
- funzione *autocovarianza* $\gamma_{t_1, t_2} = E\{[Y_{t_1} - \mu_{t_1}][Y_{t_2} - \mu_{t_2}]\}$

La funzione di autocovarianza

L'*autocovarianza* è la covarianza tra v.c. dello stesso processo stocastico spaziate di uno sfasamento temporale (*lag*) pari a $k = |t_2 - t_1|$.

- La varianza è pertanto il caso particolare per cui $t_1 = t_2$, o $k = 0$.
- La varianza e la autocovarianza sono espresse nel quadrato dell'unità di misura di Y .

L'autocovarianza può essere normalizzata ottenendo l'*autocorrelazione*:

$$\rho_{t_1, t_2} = \frac{\gamma_{t_1, t_2}}{\sigma_{t_1} \sigma_{t_2}}$$

che è un numero puro.



Processi stazionari

Un processo stocastico è *stazionario* se “la sua media e la sua variabilità non presentano cambiamenti di natura sistematica”

Più formalmente, è detto: *stazionario in senso stretto* se le distribuzioni congiunte di $(Y_{t_1}, \dots, Y_{t_n})$ e $(Y_{t_1+\tau}, \dots, Y_{t_n+\tau})$ sono uguali per ogni insieme di istanti t_1, \dots, t_n e ogni τ ; pertanto per $n = 1$ riesce

$$\mu_t = \mu$$

$$\sigma_t^2 = \sigma^2$$

e per la funzione di autocovarianza è

$$\gamma_{t_1, t_2} = E\{[Y_{t_1} - \mu_{t_1}][Y_{t_2} - \mu_{t_2}]\} = \gamma_{|t_1 - t_2|}$$

ovvero $\forall k = 0, \pm 1, \pm 2, \dots$ è

$$\gamma_{t, t+k} = E\{[Y_t - \mu_t][Y_{t+k} - \mu_{t+k}]\} = \gamma_k$$



L'autocovarianza/autocorrelazione

L'autocovarianza di un processo stazionario, indice delle relazioni lineari tra le v.c. componenti,

- è una funzione *pari* di k (i.e., $\gamma_{-k} = \gamma_k$)

La *funzione di autocorrelazione* (ACF)

$$\rho_{t,t+k} = \rho_k = \frac{\gamma_k}{\gamma_0}$$

soddisfa le:

- $\rho_0 = 1$
- $\rho_k = \rho_{-k}$
- $|\rho_k| \leq 1$

Il grafico dei valori di ρ_k , $k = 0, 1, \dots$ viene chiamato *correlogramma*.

L'autocorrelazione parziale

La correlazione tra due variabili può essere “diretta” o anche dovuta al fatto che esiste una terza variabile con cui entrambe sono correlate:

- Nel contesto delle serie storiche, buona parte della correlazione tra Y_t e Y_{t-k} può essere dovuta al fatto che entrambe sono correlate con $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$.

La *funzione di autocorrelazione parziale* (PACF) rappresenta le autocorrelazioni ai vari ordini *al netto delle variabili intermedie*, ovvero

- per ogni ordine k , P_k è l'autocorrelazione *condizionata*

$$P_k = \text{Corr}(Y_t, Y_{t-k} | Y_{t-1}, \dots, Y_{t-k+1})$$

P_k si calcola come quoziente di due determinanti di matrici in ρ_j , $j = 1, \dots, k$ (Di Fonzo e Lisi, Eq. 5.10)



Stazionarietà in senso debole

Un processo stocastico Y_t è detto *stazionario in senso debole* se

$$E(Y_t) = \mu \quad \forall t$$

$$\text{Cov}[Y_t, Y_{t+k}] = \gamma_k \quad \forall t, k$$

Per $k = 0$ risulta che anche la varianza è costante.

I processi debolmente stazionari soddisfano il *Teorema di Decomposizione di Wold*: ogni processo debolmente stazionario può essere scritto come

$$Y_t = Z_t + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$

dove Z_t è una componente deterministica, ϵ_t è *white noise*, $\psi_0 = 1$ e $\sum_{i=1}^{\infty} \psi_i^2 < +\infty$



Invertibilità

Dato un processo stocastico stazionario, è *unica la sua funzione di autocovarianza*.

Non necessariamente è vero il contrario, ovvero *non necessariamente la funzione di autocovarianza caratterizza pienamente un processo*.

- Intuizione: l'invertibilità è “la capacità di esprimere il processo in funzione dei suoi valori passati” .
- Formalmente, un processo $\{Y_t\}$ è detto *invertibile* se esiste una funzione lineare $h(\cdot)$ e un processo *white noise* $\{\epsilon_t\}$ tale che, $\forall t$, si ha:

$$Y_t = h(Y_{t-1}, Y_{t-2}, \dots) + \epsilon_t$$



Stima dei momenti di un processo stocastico

In generale, il processo stocastico generatore dei dati (DGP) di una serie storica

$$\{y_t\}_{t=1}^n$$

prodotta da un fenomeno reale è ignoto; le sue caratteristiche andranno ricercate mediante inferenza sui dati osservati, che costituiscono l'*unica realizzazione conosciuta* del processo.

In altre parole bisogna stimare i momenti in base ai dati osservati della serie storica. L'espressione

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n y_t$$

è uno stimatore corretto e consistente di μ sotto condizione di *ergodicità rispetto alla media*.



Stima dei momenti: l'autocovarianza

Uno stimatore appropriato per l'autocovarianza è

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \hat{\mu})(y_{t+k} - \hat{\mu})$$

che sotto ergodicità è uno stimatore consistente. La sua media

$$E[\hat{\gamma}_k] = \gamma_k - \frac{k}{n} \gamma_k - \frac{n-k}{n} \text{Var}[\hat{\mu}]$$

quindi esso è distorto ma la distorsione tende a 0 al divergere di n : serve “ n sufficientemente grande” e $k \ll n$ (una buona regola è $k \leq \frac{n}{4}$).

Ponendo $k = 0$ si ottiene lo stimatore della varianza γ_0 e da qui quello dell'autocorrelazione $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$.



Varianza dell'autocorrelazione: approssimazioni

Sotto opportune ipotesi, vale l'*approssimazione di Bartlett*:

$$\text{Var}[\hat{\rho}_k] \approx \frac{1}{n} \sum_{i=-\infty}^{+\infty} (\rho_i^2 + \rho_{i-k}\rho_{i+k} + 2\rho_i^2\rho_k^2 - 4\rho_i\rho_k\rho_{i-k})$$

Per $\rho_k \approx 0$, $k > q$ si può ulteriormente approssimare:

$$\text{Var}[\hat{\rho}_k] \approx \frac{1}{n} \sum_{i=-q}^q \rho_i^2 = \frac{1}{n} (1 + 2 \sum_{i=1}^q \rho_i^2)$$

L'autocorrelazione *parziale* P_k può essere stimata sostituendo $\hat{\rho}_k$ per ρ_k nella (5.10) del Di Fonzo e Lisi; per la sua varianza vale l'*approssimazione di Quenouille*: $\text{Var}[\hat{P}_k] \approx \frac{1}{n}$.

Processo white noise (WN)

Il processo *white noise* (WN) è una sequenza di v.c. $\{\epsilon_t\}$

- incorrelate
- a media e varianza costanti

Si può supporre senza perdita di generalità che la media sia nulla.

La funzione di autocorrelazione è

$$\rho_k = \begin{cases} 1, & k = 0 \\ 0, & k = \pm 1, \pm 2, \dots \end{cases}$$

La funzione di autocorrelazione parziale, data l'incorrelazione, è

$$P_k = \rho_k$$

Perciò il processo WN è debolmente stazionario.



Processo white noise - 2

Usando l'approssimazione di Bartlett,

$$\text{Var}[\hat{\rho}_k] \approx \frac{1}{n}, \forall k \neq 0$$

pertanto, asintoticamente, per l'autocorrelazione empirica di un processo WN vale la

$$\sqrt{n}\hat{\rho}_k \sim N(0, 1)$$

da cui

$$\hat{\rho}_k \sim N\left(0, \frac{1}{n}\right)$$

che si può usare per verificare l'ipotesi che i residui di un modello stimato siano incorrelati. (L'eventuale rifiuto comporta usualmente l'abbandono dell'intero modello.)

Processo a media mobile (MA)

Sia $\{\epsilon_t\}$ un WN di media 0 e varianza σ_ϵ^2 ; il *processo a media mobile* di ordine q è definito come

$$Y_t = \epsilon_t - \theta_1\epsilon_{t-1} - \dots - \theta_q\epsilon_{t-q}$$

In questo processo, indicato con $MA(q)$, la variabile Y_t è pensata come una “media” pesata di impulsi casuali presenti e passati (una “somma” in realtà perché i “pesi” non sommano necessariamente a 1).

Usando l'operatore ritardo $B : B^h Y_t = Y_{t-h}$ e indicando con

$$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$$

l'*operatore polinomiale a media mobile*, il processo MA si può scrivere come

$$Y_t = \theta(B)\epsilon_t$$



Processo a media mobile - 2

La media di un processo MA è:

$$E[Y_t] = E[\epsilon_t - \theta_1\epsilon_{t-1} - \dots - \theta_q\epsilon_{t-q}] = 0$$

L'autocovarianza:

$$\gamma_k = \begin{cases} \sigma_\epsilon^2(1 + \sum_{i=1}^q \theta_i^2), & k = 0 \\ \sigma_\epsilon^2(-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_{q-k}\theta_q), & k = 1, \dots, q \\ 0, & k > q \end{cases}$$

pertanto la funzione di autocorrelazione è:

$$\rho_k = \begin{cases} 1, & k = 0 \\ \frac{\sigma_\epsilon^2(-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_{q-k}\theta_q)}{1 + \theta_1^2 + \dots + \theta_q^2}, & k = 1, \dots, q \\ 0, & k > q \end{cases}$$



Processo a media mobile - 3

Inoltre:

- L' autocorrelazione parziale P_k tende ad annullarsi al divergere di k .
- Poiché media, varianza e funzione di autocovarianza non dipendono dal tempo, un processo MA è *sempre stazionario*.
- La condizione di invertibilità si ottiene dall'*equazione caratteristica* $\theta(B) = 0$: si dimostra che un processo MA è invertibile se

$$|B_i| < 1, i = 1, \dots, q$$

tutte le radici dell'equazione caratteristica sono, in modulo, maggiori di uno.



Processo autoregressivo (AR)

Sia $\{\epsilon_t\}$ un WN di media 0 e varianza σ_ϵ^2 ; il *processo autoregressivo* di ordine p , indicato con $AR(p)$, è definito come

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t$$

La v.c. Y_t è pensata come il risultato di una somma pesata di valori passati e di uno *shock* casuale contemporaneo. Indicando con

$$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$$

l'operatore polinomiale autoregressivo, un processo AR può essere scritto come

$$\phi(B)Y_t = \epsilon_t$$



Processo autoregressivo - 2

La media di un processo $AR(p)$ è pari a:

$$\mu = E[Y_t] = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}$$

quindi $E[Y_t] = 0 \iff \phi_0 = 0$, e $\phi_0 = \mu(1 - \phi_1 - \dots - \phi_p)$ da cui il modello si può scrivere in scarti dalla media come:

$$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \dots + \phi_p(Y_{t-p} - \mu) + \epsilon_t$$

La funzione di autocovarianza:

$$\gamma_k = \begin{cases} \phi_1\gamma_1 + \dots + \phi_p\gamma_p + \sigma_\epsilon^2, & k = 0 \\ \phi_1\gamma_{k-1} + \dots + \phi_p\gamma_{k-p}, & k > 0 \end{cases}$$

Processo autoregressivo - 3

La varianza può essere espressa come:

$$\gamma_0 = \frac{\sigma_\epsilon^2}{1 - \phi_1\rho_1 - \dots - \phi_p\rho_p}$$

Per la funzione di autocorrelazione:

$$\rho_k = \phi_1\rho_{k-1} + \dots + \phi_p\rho_{k-p}, k = 0$$

che, scritta per esteso e considerato che $\rho_0 = 1$ e $\rho_k = \rho_{-k}$, dà luogo alle *Equazioni di Yule-Walker*:

$$\begin{cases} \rho_1 = \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \dots + \phi_p\rho_{p-1} \\ \rho_2 = \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \dots + \phi_p\rho_{p-2} \\ \dots \\ \rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \dots + \phi_p \end{cases}$$

Processo autoregressivo - 4

Le equazioni di Yule-Walker consentono di passare dalla funzione di autocorrelazione ai parametri del processo $AR(p)$.

Inoltre si può mostrare che:

- in un processo AR stazionario, l'ACF ρ_k tende ad annullarsi al divergere di k
- la funzione di autocorrelazione parziale P_k è diversa da 0 per $k \leq p$, poi si annulla per $k > p$
- il processo AR è sempre invertibile (Y_t è già espressa in funzione dei suoi valori passati)
- un processo AR è stazionario se tutte le radici dell'equazione caratteristica $\phi(B) = 0$ sono, in modulo, maggiori di 1

Processo autoregressivo a media mobile (ARMA)

Sia, al solito, $\{\epsilon_t\}$ un WN di media 0 e varianza σ_ϵ^2 . Il *processo autoregressivo a media mobile* di ordine p, q , indicato con $\text{ARMA}(p, q)$, è definito come

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$$

Il processo è la combinazione di una parte AR e di una MA.

Indicando con $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ l'operatore polinomiale autoregressivo e con $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ l'*operatore polinomiale a media mobile*, il processo ARMA si può scrivere in forma più compatta come

$$\phi(B)Y_t = \phi_0 + \theta(B)\epsilon_t$$

Questi processi sono considerati la classe standard per serie stazionarie.



Processo ARMA - 2

Un processo ARMA(p, q) è

- stazionario se tutte le radici dell'equazione caratteristica $\phi(B) = 0$ sono, in modulo, maggiori di 1
- invertibile se tutte le radici dell'equazione caratteristica $\theta(B) = 0$ sono, in modulo, maggiori di 1.

La funzione di autocovarianza:

$$\gamma_k = \begin{cases} \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + E[Y_{t-k} \epsilon_t] - \theta_1 E[Y_{t-k} \epsilon_{t-1}] - \\ \quad \dots - \theta_q E[Y_{t-k} \epsilon_{t-q}], & k = 0, 1, \dots, q \\ \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}, & k \geq q + 1 \end{cases}$$

Fino al ritardo q la funzione di autocovarianza dipende da entrambe le “parti” del processo; per $k > q$ solo da quella autoregressiva.



Processo ARMA - 3

Per la funzione di autocorrelazione:

$$\rho_k = \begin{cases} \frac{\gamma_k}{\gamma_0}, & k = 0, 1, \dots, q \\ \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}, & k \geq q + 1 \end{cases}$$

In generale, le ACF e PACF rispecchiano i comportamenti di AR e MA:

- Per $k > q$, la ACF di un processo ARMA(p, q) si comporta come quella di un AR(p), ovvero tende ad annullarsi al divergere di k
- Per $k > p$, la PACF tende esponenzialmente a 0 come per un processo MA(q)

Processo ARMA (p, q) come AR(∞)

Si noti che un processo ARMA(p, q) stazionario ed invertibile si può sempre esprimere come un AR(∞) del tipo

$$(1 - \pi_1 B - \pi_2 B^2 - \dots) Y_t = \epsilon_t$$

Infatti,

$$\phi(B) Y_t = \theta(B) \epsilon_t$$

è equivalente a:

$$\frac{\phi(B)}{\theta(B)} Y_t = \epsilon_t$$

che è un AR del tipo $\pi(B) Y_t = \epsilon_t$ con

$$\pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \dots) = \phi(B) \theta(B)^{-1}$$

Processo ARMA (p, q) come $MA(\infty)$

Analogamente, un ARMA (p, q) stazionario ed invertibile si può sempre esprimere come un $MA(\infty)$ del tipo

$$Y_t = \psi(B)\epsilon_t$$

con

$$\psi(B) = (1 - \psi_1 B - \psi_2^2 B^2 - \dots) = \phi^{-1}(B)\theta(B)$$

e i coefficienti, rispettivamente, autoregressivi π_i e a media mobile ψ_i si ottengono uguagliando i coefficienti delle stesse potenze di B .

Processi non stazionari

La non stazionarietà può dipendere dal fatto che non sono costanti

- i momenti primi (*non stazionarietà in media*)
- i momenti secondi (*non stazionarietà in (co)varianza*)

Si consideri un WN più un trend deterministico lineare:

$$Y_t = \beta_0 + \beta_1 t + \epsilon_t$$

la media è $E[Y_t] = \beta_0 + \beta_1 t$ che non è costante; tuttavia

$$(I - B)Y_t = Y_t - Y_{t-1}$$

è un MA(1) e quindi è stazionario: infatti

$$\begin{aligned}(I - B)Y_t &= \beta_0 + \beta_1 t + \epsilon_t - (\beta_0 + \beta_1(t - 1) + \epsilon_{t-1}) \\ &= \beta_1 + \epsilon_t - \epsilon_{t-1}\end{aligned}$$



Processi non stazionari omogenei e non

Un processo si dice *omogeneo di grado d* se può essere ridotto alla stazionarietà con d differenziazioni successive.

Può darsi che la differenziazione non risolva. E' il caso dei processi non stazionari in varianza, dove per esempio la varianza può essere funzione della media:

$$\text{Var}[Y_t] = cf(\mu_t)$$

Per esempio:

$$\text{Var}[Y_t] = c^2 \mu^2$$

in questo caso la varianza può essere stabilizzata usando il logaritmo. In generale, si può usare la *trasformazione di Box-Cox*

$$T(Y_t) = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln Y_t, & \lambda = 0 \end{cases}$$

Processo random walk

Sia $\{\epsilon_t\}$ un WN di media 0 e varianza σ_ϵ^2 . Il *processo random walk* (passeggiata casuale) è definito come

$$Y_t = Y_{t-1} + \epsilon_t, \quad Y_0 = \mu$$

Essendo $Y_t = \mu + \sum_{j=1}^t \epsilon_j$, si ha:

$$\begin{cases} E[Y_t] = \mu \\ \text{Var}[Y_t] = t\sigma_\epsilon^2 \end{cases}$$

Il processo random walk è un AR(1) con $\phi = 1$ che pertanto non rispetta la condizione di stazionarietà $|\phi| < 1$. Si vede anche che la varianza è funzione crescente del tempo.

Processo random walk - 2

La funzione di autocorrelazione di un random walk:

$$\rho_k = \sqrt{\frac{t}{t-k}}, \quad k = 0, 1, 2, \dots$$

pertanto al crescere di t , e per ogni k , $\rho_k \rightarrow 1$ (si usa dire che il processo ha *memoria infinita*).

La funzione di autocorrelazione parziale:

$$P_k = \begin{cases} 1 & k = 1 \\ 0 & k > 1 \end{cases}$$

La differenza prima $(I - B)Y_t$ è un processo white noise, che come tale è stazionario. Il random walk è pertanto *omogeneo di grado 1*.

Processo ARIMA

Il processo ARMA può essere esteso ai processi non stazionari omogenei di grado d . Sia, al solito, $\{\epsilon_t\}$ un WN di media 0 e varianza σ_ϵ^2 .

Indicata con X_t la d -esima differenza di Y_t , $X_t = (1 - B)^d Y_t$, Y_t si dirà *processo autoregressivo integrato a media mobile* di ordine p, d, q (ARIMA(p, d, q)) se X_t è ARMA(p, q). Ovvero:

$$X_t = (1 - B)^d Y_t$$

$$X_t = \phi_0 + \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

In termini di Y e operatori polinomiali,

$$\phi(B)(1 - B)^d Y_t = \phi_0 + \theta(B)\epsilon_t$$

Tutti gli AR/MA/ARMA sono casi particolari. Il RW è ARIMA(0, 1, 0) e pertanto si indica anche con I(1).

Processi stagionali

I modelli ARIMA possono essere estesi alla stagionalità di carattere *stocastico* (ancora Box e Jenkins 1976), introducendo un tipo di comportamento periodico che può anche essere non stazionario. Tali processi sono detti *ARIMA stagionali*, o SARIMA.

- Si noti che qui non interessa separare le componenti ma modellarle congiuntamente.

Si modellano due tipi di correlazioni:

- tra osservazioni consecutive (correlazione “orizzontale”)
- tra osservazioni che distano tra loro di un multiplo del periodo (corr. “verticale”)

(tabella)



Processi stagionali - i dati Airlines

(in R, AirPassengers)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	209	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306
1957	315	301	356	348	355	422	465	467	404	347	305	336
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	508	461	390	432

Processi SARIMA

Si assume dunque che Y_t segua un processo ARIMA

$$\phi(B)(1 - B)^d Y_t = \phi_0 + \theta(B)b_t$$

dove però b_t non è WN ma contiene ancora una componente stagionale:

$$\Phi(B^S)(1 - B^S)^D b_t = \Theta(B^S)\epsilon_t$$

e B^S è l'operatore ritardo *stagionale* e quindi

$$\Phi(B^S) = (1 - \phi_1 B^S - \phi_2 B^{2S} + \dots + \phi_P B^{PS})$$

$$\Theta(B^S) = (1 - \theta_1 B^S + \theta_2 B^{2S} + \dots + \theta_Q B^{QS})$$

(dove S è il periodo stagionale, es. $S = 12$ per dati mensili) i corrispondenti operatori polinomiali autoregressivo e a media mobile.



Processi SARIMA - 2

Combinando le due strutture di dipendenza (sugli anni e sulle stagioni) si ottiene il modello $SARIMA(p, d, q) \times (P, D, Q)_S$

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D Y_t = \phi_0 + \theta(B)\Theta(B^S)\epsilon_t$$

Applicando gli operatori differenza in sequenza,

$$Z_t = (1-B)^d(1-B^S)^D Y_t$$

segue un modello $SARMA(p, 0, q) \times (P, 0, Q)_S$.

Moltiplicando tra loro gli operatori polinomiali stagionali e non, è facile (!) vedere che il modello $SARMA(p, 0, q) \times (P, 0, Q)_S$ può essere scritto come un $ARMA(p + PS, q + QS)$.

Le condizioni di stazionarietà e invertibilità possono essere desunte da quest'ultima forma.

Processi SARIMA - 3

Il modello SARMA($p, 0, q$) \times ($P, 0, Q$) $_S$ sarà

- stazionario se le radici di $\phi(B)\Phi(B^S) = 0$ sono tutte $|\cdot| > 1$
- invertibile se le radici di $\theta(B)\Theta(B^S) = 0$ sono tutte $|\cdot| > 1$

Anche i processi stagionali possono essere caratterizzati tramite ACF e PACF: es. il processo ARIMA($0, 0, 0$) \times ($1, 0, 0$) $_S$

$$(1 - \Phi_1 B^S) Y_t = \epsilon_t$$

ha funzione di autocorrelazione semplice

$$\rho_k = \begin{cases} \Phi_1^{\frac{k}{S}}, & k = 0, S, 2S, 3S, \dots \\ 0, & k \neq 0, S, 2S, 3S, \dots \end{cases}$$

e funzione di autocorrelazione parziale

$$P_k = \begin{cases} \Phi_1, & k = S \\ 0, & k \neq S \end{cases}$$



La procedura di Box e Jenkins

Cercasi processo stocastico ARIMA che possa plausibilmente aver generato i dati che osserviamo.

Box e Jenkins (1976) propongono una procedura standard (un algoritmo, se volete) per arrivarci:

- Identificazione
- Stima dei parametri
- Controllo diagnostico
- (repeat?)

Identificazione

Specificare il modello significa identificare l'ordine p, d, q di ogni componente. Gli strumenti di indagine sono le funzioni ACF e PACF campionarie.

L'idea di base è di *riconoscere nella struttura della (P)ACF campionaria quella di una (P)ACF teorica*, sulla base delle seguenti considerazioni:

- a la ACF e la PACF di un ARMA si annullano a velocità esponenziale in k . Se restano “grandi troppo a lungo” si sospetta non-stazionarietà:
 - ▶ se c'è un trend deterministico si detrendizza
 - ▶ se c'è un trend stocastico si differenzia
- b per un MA(q) l'ACF *teorica* si annulla per $k > q$ (= nel campione tenderà ad assumere valori non significativamente diversi da 0), la PACF tende a zero esponenzialmente
- c per un AR(p) viceversa: la ACF tende a zero esponenzialmente, la PACF si annulla per $k > p$

Quando la situazione non è chiaramente b o c, si assume un ARMA.



Identificazione - 2

Se la situazione non è chiara, un modello ARMA ha il vantaggio di essere parsimonioso in termini di p e q . Due considerazioni:

- il *fit in-sample* del modello non farà che aumentare con il numero dei parametri
- un modello *sovraspecificato* (=con p , q troppo grandi) funzionerà spesso peggio in previsione di uno parsimonioso.

Per decidere quali ordini assegnare a p e q si possono usare due criteri basati sulla massima verosimiglianza del modello, che

- premiano il “fit”
- penalizzano il numero dei parametri

Criteri per la specificazione

Il criterio di Akaike:

$$AIC(k) = -\frac{2}{n}(\log L(\delta) - k)$$

Il criterio di Schwarz:

$$SC(k) = -\frac{2\log L(\delta)}{n} + k\frac{\log n}{n}$$

Tra i due AIC tende alla sovrapparametrizzazione.

Alternativamente, si può stimare un modello “abbondante” e ridurlo dal generale al particolare usando i *test di significatività* dei parametri.

Stima dei parametri

Due metodi principali:

- minimi quadrati non lineari (NLS)
- massima verosimiglianza (ML)

I NLS non richiedono ipotesi distributive; la ML invece richiede ipotesi sulla distribuzione congiunta del processo, ma è lo stimatore più efficiente.

Alcuni modelli ammettono una stima in forma chiusa (AR) ma in generale la stima richiede procedure numeriche (ottimizzazione)

Stima di massima verosimiglianza

La ML parte dalla *funzione di verosimiglianza*

$$L(\delta) = f(y_1, \dots, y_n | \delta)$$

la stima di ML è quel valore $\hat{\delta}$ tale che

$$L(\hat{\delta}) = \max_{\delta} L(\delta)$$

Usare la distribuzione congiunta è proibitivo; pertanto si usa scomporre

$$L(\delta) = f(y_1, \dots, y_m | \delta) \prod_{t=m+1}^n f(y_t | I_{t-1}; \delta)$$

la congiunta fino a m e il prodotto delle densità *condizionate* da $m+1$ a n .
Per ragioni computazionali, si usa massimizzare la *log-verosimiglianza*: il che è equivalente (il logaritmo è una trasformazione monotona).



Controllo diagnostico

- analisi grafiche
- autocorrelazione dei residui
- test di tipo *portmanteau*
- verifica di casualità dei residui
- test di normalità dei residui

(plot: dati AirPassengers)

Autocorrelazione dei residui

(ACF/PACF plot: dati AirPassengers)

Test portmanteau

Un test *portmanteau* ha una nulla ben specificata e una alternativa specificata in modo generico; pertanto ha potenza contro un range ampio di alternative similari (può avvertire di deviazioni dalla nulla di vario tipo).

La statistica di Q di Ljung e Box

$$Q(m) = n(n+2) \sum_{k=1}^m \frac{1}{n-k} \hat{\rho}_k^2$$

verifica l'assenza di autocorrelazione in modo "complessivo" fino all' m -esimo ritardo. Sotto H_0 si distribuisce come χ_{m-p-q}^2 .

Attenti ai gradi di libertà della χ in Box.test



Verifica di casualità dei residui

I residui possono essere sottoposti a una serie di test che ne verifichino la “casualità” (si veda per esempio il test dei punti di svolta)

Test di normalità

Nel caso i residui siano normalmente distribuiti, la loro incorrelazione implica l'indipendenza.

In un processo gaussiano a componenti incorrelate, detto m_k il k -esimo momento campionario centrato, il coefficiente di asimmetria

$$\hat{S} = \frac{m_3}{m_2^{3/2}}$$

con, sotto normalità e incorrelazione, $\sqrt{\frac{n}{6}} \hat{S} \sim N(0, 1)$.

Il coefficiente di curtosi

$$\hat{K} = \frac{m_4}{m_2^2}$$

con, idem, $\sqrt{\frac{n}{24}} (\hat{K} - 3) \sim N(0, 1)$ da cui la statistica di Jarque e Bera:

$$JB = \hat{S}^2 + \hat{K}^2 \sim \chi_2^2$$

sotto l'ipotesi nulla.



Cenni di teoria della previsione ottima

La *previsione* come stima di un valore caratteristico di una v.c. tramite un prefissato *insieme di informazione* al tempo t (I_t). Nel nostro caso,

$$I_t = \{Y_t, Y_{t-1}, \dots, Y_0, Y_{-1}, \dots\}$$

- Fare previsioni significa fare affermazioni sulla v.c. Y_{t+k} *condizionatamente ad* I_t cioè al passato del processo.
- La variabile Y_{t+k} è detta *predicibile* rispetto a I_t se $F(Y_{t+k}|I_t) \neq F(Y_{t+k})$; e viceversa.

Il *previsore* $\hat{Y}_{t+k} = g(I_t)$ è una funzione del passato.

Indicheremo con

- \hat{Y}_{t+k} il *previsore*
- \hat{y}_{t+k} la *previsione*.



Il previsore ottimo

La funzione g è scelta in modo da minimizzare una funzione di perdita, che solitamente è connessa all'errore di previsione:

$$e_{t+k} = Y_{t+k} - \hat{Y}_{t+k} = Y_{t+k} - g(I_t)$$

Inoltre, è ragionevole richiedere che l'errore di previsione atteso sia nullo e che la funzione scelta minimizzi, a priori, la varianza dell'errore:

$$E(e_{t+1}) = E(Y_{t+k} - g(I_t)) = 0 \quad \forall t$$

$$E(e_{t+1}^2) = E[(Y_{t+k} - g(I_t))^2] = \min_g$$

La quantità $E(e_{t+1}^2)$ (errore quadratico medio, EQM/MSQ) è la funzione di perdita di riferimento.

Si dimostra che la funzione ottima è il *valore atteso condizionato*

$$\hat{Y}_{t+k} = E(Y_{t+k} | I_t)$$

la cui espressione, al di fuori del caso lineare, non è unica.

La previsione con modelli ARIMA

Ipotizzando di avere la serie $\{y_t\}_{t=1}^n$ e di aver stimato un modello ARIMA(p, d, q) che farà da previsore; il calcolo di \hat{y}_{t+k} si basa sulle ipotesi:

- corretta identificazione del modello
- conoscenza dei parametri ϕ, θ

pertanto il modello stimato fornirà un'approssimazione al previsore ottimale.

Nella pratica, in ogni istante t si porrà $E[Y_j|I_t] = y_j$ per ogni valore passato ($j < t$) e perciò i corrispondenti errori di previsione $e_j = 0 \forall j < t$. L'errore di previsione futuro $e_{t+k} = y_{t+k} - \hat{y}_{t+k}$ sarà una v.c. Al crescere di k , per modelli ARMA stazionari, riesce

$$\lim_{k \rightarrow \infty} E[Y_{t+k}|I_t] = E[Y_t]$$

$$\lim_{k \rightarrow \infty} \text{Var}[e_{t+k}] = \text{Var}[Y_t]$$



Previsione intervallare

Oltre alla previsione *puntuale* è interessante effettuare una previsione *intervallare*, ovvero trovare un intervallo $[A, B]$:

$$P(A \leq Y_{n-k} \leq B) = 1 - \alpha$$

Se l'errore di previsione ha media nulla, allora la previsione intervallare in $t + k$ sarà centrata su y_{t+k} e la sua ampiezza dipenderà dalla varianza dell'errore di previsione.

Dalla normalità di ϵ_t discende quella dell'errore di previsione e da questa quella del previsore: se i parametri del modello sono noti, è

$$\hat{Y}_{t-k} \sim N(y_{t+k}, \text{Var}[e_{t+k}])$$

e gli estremi dell'intervallo:

$$\hat{y}_{t+k} \pm z_{1-\frac{\alpha}{2}} \sqrt{\text{Var}[e_{t+k}]}$$

(ma nella pratica i parametri vanno stimati e l'intervallo “si allarga”)



Livellamento esponenziale e modelli ARIMA

Si consideri il lisciamento esponenziale semplice:

$$\begin{aligned}\hat{y}_{n+1} &= y_n - \delta(y_n - \hat{y}_n) \\ &= y_n - \delta e_n\end{aligned}$$

esso è il previsore ottimale se minimizza l'EQM e se la successione degli errori di previsione è incorrelata.

Siccome il previsore che minimizza l'EQM è il valore atteso condizionato, $y_{n+1} = \hat{y}_{n+1} + e_{n+1}$ ovvero $\hat{y}_{n+1} = y_{n+1} - e_{n+1}$ quindi

$$y_n - \delta e_n = y_{n+1} - e_{n+1}$$

e dunque $y_{n+1} = y_n - \delta e_n + e_{n+1}$ ovvero

$$(1 - B)y_{n+1} = (1 - \delta B)e_{n+1}$$

che ha una struttura di tipo ARIMA(0, 1, 1) con $\theta = \delta$.



Livellamento esponenziale e modelli ARIMA - 2

Questo significa che

- il previsore ottenuto con il lisciamento esponenziale è ottimale solo se la serie storica è stata generata da un processo ARIMA(0, 1, 1) con $0 < \theta < 1$.

Inoltre, si può mostrare che:

- il previsore basato sul metodo di Holt e Winters è ottimale solo se il processo generatore è un ARIMA(0, 2, 2)
- la corrispondente versione stagionale è ottimale solo per dati presi da un ARIMA(0, 2, 2,) \times (0, 1, 1)_S

I modelli perequativi sono pertanto dei casi particolari dei modelli ARIMA, che rappresentano una classe più generale.



Processi trend-stazionari e radici unitarie

Consideriamo i due diversi tipi incontrati di non-stazionarietà:

- stazionarietà attorno a una funzione deterministica, p. es. un trend lineare:

$$Y_t = \beta_0 + \beta_1 t + \epsilon_t$$

- processo a *radice unitaria*

$$(1 - B)Y_t = \beta + \psi(B)\epsilon_t$$

Un classico esempio del secondo è il *random walk* con *drift*:

$Y_t = \beta + Y_{t-1} + \epsilon_t$ che viene detto anche *integrato di ordine 1* o $I(1)$.

In genere, i processi del secondo tipo vengono definiti *a trend stocastico*.

Trend deterministico vs. stocastico: la previsione

Si può dimostrare che:

- il processo a trend deterministico converge (in media quadratica) al trend al crescere dell'orizzonte temporale k :

$$\lim_{k \rightarrow \infty} E[\hat{Y}_{t+k} - \beta_0 - \beta_1(t+k)]^2 = 0$$

- la previsione (in t) del processo $I(1)$ è data dalla $\hat{y}_{t+k} = \beta k + y_t$

dunque con entrambe le specificazioni la previsione \hat{Y}_{t+k} è una funzione lineare dell'orizzonte di previsione k con coefficiente angolare β : la differenza sta *nell'intercetta*.

- Per un processo trend-stazionario, la previsione è una retta la cui intercetta è la stessa indipendentemente dal valore di y_t (il processo è *mean-reverting* attorno al trend deterministico).
- Per un processo *unit root*, l'intercetta cambia a ogni nuova osservazione y_t (*perfetta memoria* delle innovazioni del passato)

(Es. il PIL dell'Italia dopo il Covid: ritorno al trend di lungo periodo o memoria dello shock?)



Il test di radici unitarie di Dickey e Fuller

Si consideri il processo AR(1) $Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t$ con $\{\epsilon_t\}$ white noise. Se $|\phi_1| < 1$ il processo è stazionario. Escluso il caso esplosivo, confrontiamo dunque

$$H_0 : \phi_1 = 1$$

$$H_A : \phi_1 < 1$$

Il problema è che sotto H_0 la statistica $t = \frac{\hat{\phi}_1}{SE(\hat{\phi}_1)}$

- non ha la distribuzione standard
- dipende anche dalle ipotesi sulla parte deterministica

Dickey e Fuller hanno tabulato i valori critici della distribuzione nei casi:

- trend deterministico e costante
- solo costante
- né trend deterministico né costante

e per varie numerosità campionarie. (Oggi si usa la versione ADF del test)



The end

