



FRANCESCO FUSCO

ANALISI CRITICA DEL MODELLO S.I.R.

Nel mondo ci sono state,
in egual numero,
pestilenze e guerre; e
tuttavia pestilenze e
guerre colgono gli uomini
sempre impreparati.

(Albert Camus)

INTRODUZIONE

In questa lezione impareremo a:

- Studiare il modello SIR sulle pandemie
- Comprendere alcuni concetti chiave del fenomeno
- Stabilire quali parametri descrivono la diffusione del virus nel tempo nella popolazione
- Caratterizzare matematicamente una pandemia reale
- Prevedere l'andamento nel futuro, attraverso l'impiego di un metodo speditivo del calcolo $R(t)$
- Regolare l'indice $R(t)$, tramite interventi sociali per mitigare l'impatto sul sistema sanitario

A CHE SERVE UN MODELLO PREDITTIVO

- ❑ Un **modello predittivo** è utile se descrive abbastanza bene l'evolversi di un fenomeno nel futuro
- ❑ In una pandemia, esso deve indicare come varia il numero di infetti nel tempo, permettendo di capire come si diffonde il virus e come intervenire per modificarne il corso o prepararsi per tempo a gestire le conseguenze (es. ricoveri ospedalieri).
- ❑ Un modello matematico deve poter descrivere come si trasmette un virus, attraverso delle funzioni che sono tipiche per quel virus e per quella popolazione.
- ❑ Modellare matematicamente un fenomeno significa definire: un sistema di **equazioni (differenziali)**, dei **parametri caratteristici** e alcune **condizioni al contorno**.
- ❑ Date le equazioni del modello possiamo fare delle simulazioni provando ad apportare delle modifiche ai parametri e rispondere alle domande del tipo **WHAT...IF** e dare istruzioni agli addetti ai lavori per ridurre il rischio.

IL MODELLO MATEMATICO SIR

Quando si diffonde un nuovo virus per il quale non ci sono terapie o profilassi, gli epidemiologi sono soliti applicare il modello predittivo **S.I.R.**, basato su un insieme di equazioni differenziali, regolare da due parametri, che descrivono le variazioni nel tempo di tre gruppi di individui in cui è partizionata la popolazione.

Secondo il modello **SIR** la variabile indipendente è il tempo t , misurato in giorni, mentre la variabile dipendente è il numero di persone appartenente ad uno dei 3 gruppi.

L'insieme di variabili dipendenti conta le persone in ciascuno gruppo, in funzione del tempo:

$S = S(t)$ è il numero di individui suscettibili, capaci di infettarsi con il virus

$I = I(t)$ è il numero di individui infetti e contagiosi, che trasmettono il virus

$R = R(t)$ è il numero di individui recuperati, che non possono più infettarsi

Il modello matematico SIR descrive l'andamento delle tre variabili dipendenti, con delle funzioni matematiche parametriche e, date delle condizioni al contorno, consente di prevedere il valore di ciascuna ad un tempo t futuro.

Il paradigma del gregge del modello SIR



IL GREGGE STAZIONARIO

Il modello SIR presuppone che la popolazione sia formata da una mescolanza omogenea degli individui appartenenti ai tre gruppi, con una trasmissione del virus che avviene principalmente per diffusione a catena, legata al «contatto ravvicinato» tra due individui

Inoltre il SIR è a somma costante, non tenendo conto della variazione della popolazione per varie cause (morte).

LE EQUAZIONI DELLA PANDEMIA

Poiché il modello SIR studia la variazione nel tempo del numero di individui nei tre gruppi in cui è divisibile una popolazione sotto pandemia, è opportuno rappresentare la dimensione di questi gruppi con la quota percentuale rispetto alla popolazione totale costante N , in modo da rendere il modello indipendente dal valore di N .

$s(t) = S(t) / N$, la frazione suscettibile della popolazione,

$i(t) = I(t) / N$, la frazione infetta della popolazione

$r(t) = R(t) / N$, la frazione recuperata della popolazione

E' interessante notare che vale

$$s(t) + i(t) + r(t) = 1$$

in qualsiasi istante temporale t , cioè ogni individuo appartiene necessariamente ad uno dei tre gruppi e la popolazione non cambia nel tempo di analisi

La popolazione totale è supposta costante durante l'intervallo temporale di studio, è ed formata da persone sostanzialmente statiche che infettano solo nel loro intorno (*effetto gregge*)

LE IPOTESI DEL MODELLO

Suscettibili, Infetti, Recuperati

- ❑ IP1) Nessuno individuo viene aggiunto o sottratto al gruppo dei suscettibili. L'unico modo in cui un individuo lascia il gruppo suscettibile è diventare infetto.
- ❑ IP2) il tasso del numero di suscettibili, dipende dal loro numero, dal numero degli infetti e dalla quantità di contatti tra suscettibili e infetti. Se ogni infetto ha b contatti giornalieri capaci di trasmettere il virus, con una mescolanza omogenea della popolazione, in media, ogni infetto genera $b s(t)$ nuovi infetti al giorno
- ❑ IP3) il numero dei possibili contagi giornalieri dipende linearmente dalla frazione di infetti nella popolazione $i(t)$
- ❑ IP4) una frazione k degli infetti recupera dopo m giorni, quindi in media $1/m$ della popolazione infetta si recupera (non è più infetta contagiosa) al giornalmente e quindi gli infetti diminuiscono al tasso $k i(t)$.

Le leggi di variazione del modello SIR

Il tasso di variazione (negativa) del numero dei suscettibili è proporzionale al numero degli stessi, al numero degli infetti corrente e al numero medio di contagi giornalieri, efficaci nella trasmissione

Il tasso di variazione del numero degli infetti è proporzionale al tasso di diminuzione dei suscettibili meno il tasso di variazione in aumento dei ricoverati

Il tasso di variazione dei ricoverati è proporzionale al tasso di guarigione giornaliera degli infetti

Il modello è un sistema parametrico non lineare, per la mutua interazione tra le variabili dipendenti suscettibili-infetti.

Il tasso si misura ogni singolo giorno.

I parametri b e k

b = al numero medio di contagi giornalieri da parte di un infetto contagioso, cioè quante persone riesce ad infettare un individuo quando è contagioso. Chiaramente questo parametro dipende dal comportamento degli individui suscettibili ed infetti nel gruppo. Più piccolo è b minore è il numero di nuovi infetti giornalieri

k = al tasso di guarigione giornaliera degli infetti, che è l'inverso della durata media di contagiosità data dalla somma del periodo di incubazione (asintomatico) e quello della malattia (sintomatico)
In media questo numero dipende dal tipo di virus e quindi è noto

Il rapporto $b/k=R_0$ è il parametro che regola l'andamento della pandemia

Come possiamo stimare il numero degli infetti?

E' molto semplice

Il numero dei nuovi infetti ogni giorno è dato dal numero di persone che gli infetti di ieri hanno contagiato meno il numero di persone infette che sono guarite, ma la probabilità di contagiare è data dal parametro b per il numero di persone a rischio contagio (suscettibili)

Il numero dei suscettibili diminuisce della quantità di infettati

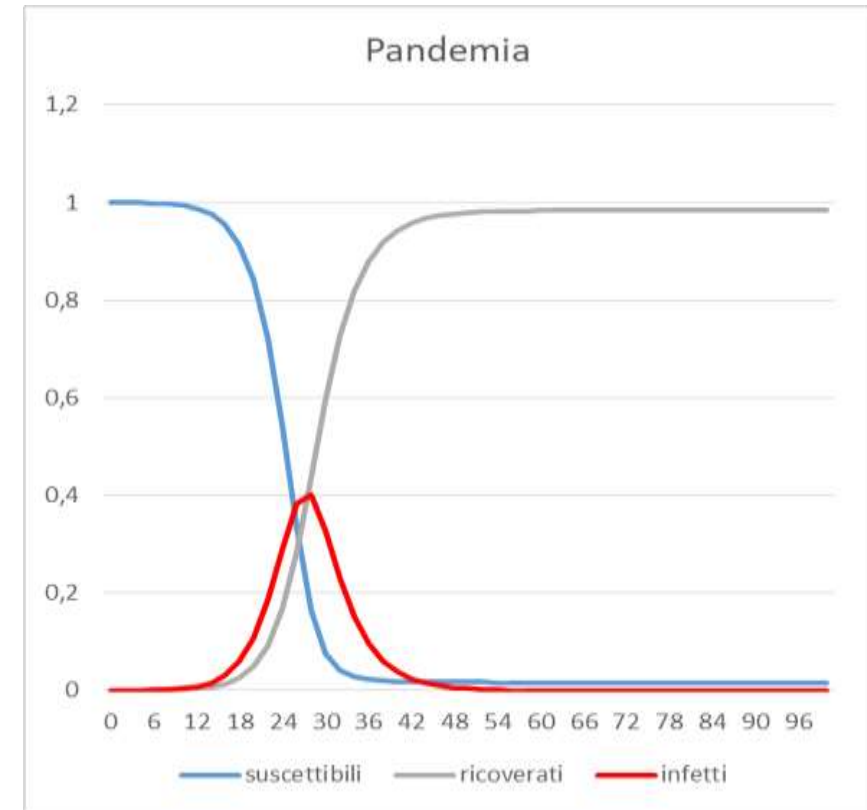
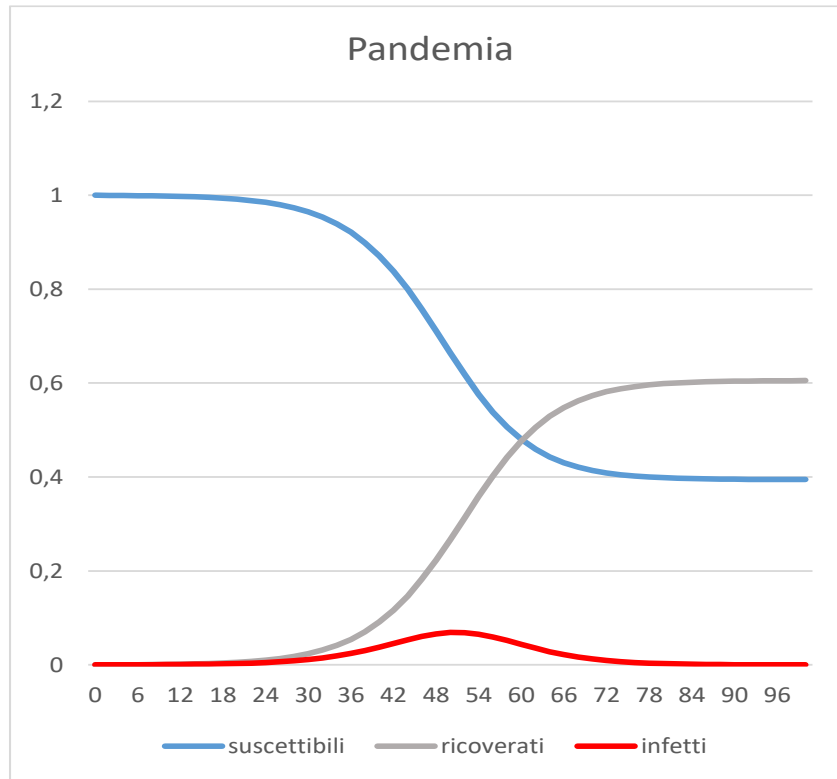
Il numero dei ricoverati aumenta della quantità dei guariti

$$s_n = s_{n-1} - b s_{n-1} i_{n-1}$$

$$i_n = i_{n-1} + (b s_{n-1} i_{n-1} - k i_{n-1})$$

$$r_n = r_{n-1} + k i_{n-1}$$

Alcuni esempi del modello SIR



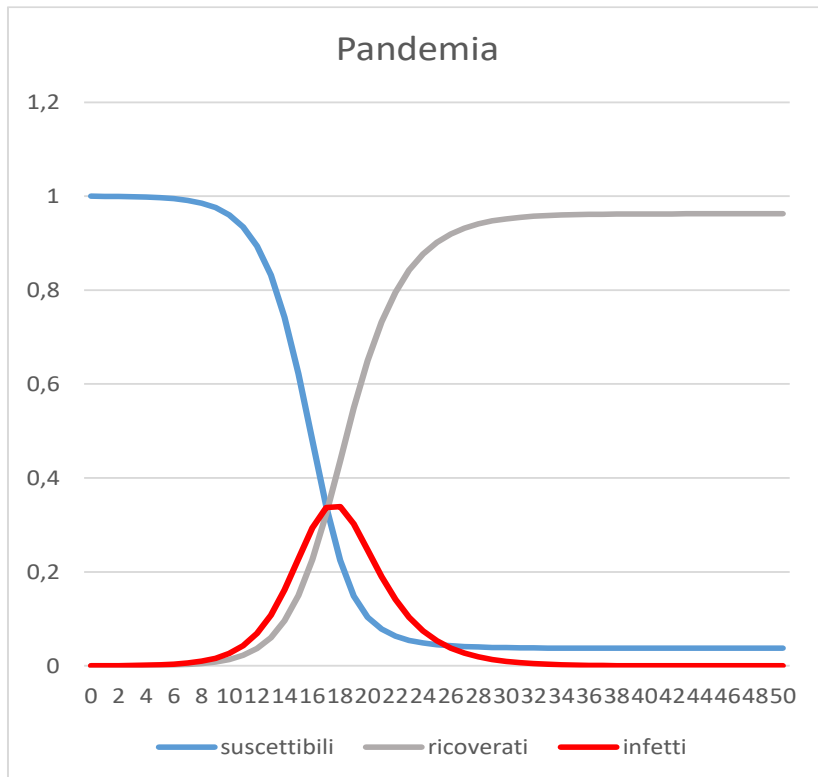
b	0,5	k	0,333333333	tasso contagio	2,0
popolazione	60000	deltaT	2	durata contagio	3,0

b	0,666667	k	0,2	tasso contagio	1,5
popolazione	60000	deltaT	2	durata contagio	5,0

Precisazioni sull'approssimazione del metodo

Data l'approssimazione numerica nella soluzione delle equazioni differenziali, con il metodo di Eulero, è importante stabilire la corretta dimensione di Δt per avere errori contenuti.

Poiché la dinamica del sistema SIR è misurata quotidianamente non ha significato andare sotto tale unità, per cui consigliamo di porre il minimo di $\Delta t=1$ e il massimo di $k=1$.



Il grafico a sinistra è elaborato con $b=1$, $k=1/3$ e $\Delta t=1$, mostra due aspetti importanti

- Il massimo degli infetti dipende da b e k
- Esiste una frazione minima dei suscettibili che non possono essere infettati quando non ci sono praticamente più infetti nella popolazione (**immunità di gregge effettiva**)

Immunità di gregge naturale

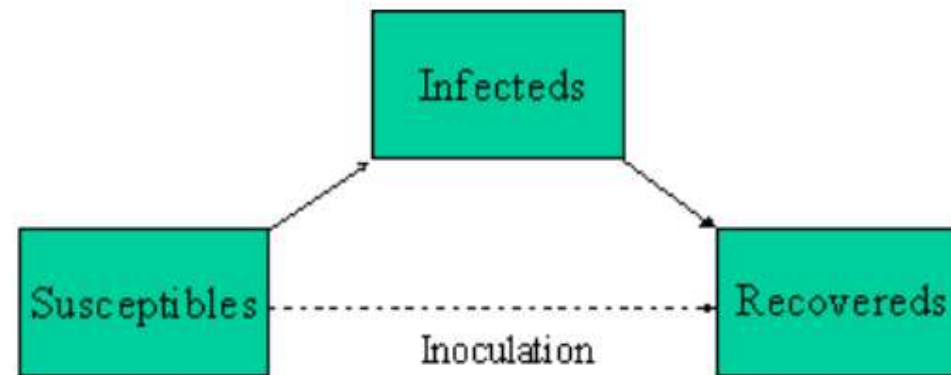
Avrete sentito parlate di immunità di gregge, spero senza che vi siate offesi, cioè del fatto che quando il numero dei suscettibili s è molto basso rispetto ai ricoverati praticamente diventa poco probabile infettarsi, perché la probabilità di contagio data da $b \times s$ è piccola.

Ma via via che tale probabilità diminuisce con il tempo, si arriva ad un punto che la probabilità è talmente bassa che il numero nuovo di infetti giornalieri è minore di 1!!!

E' quando questo succede praticamente non c'è più un infetto aggiuntivo e dobbiamo solo aspettare che gli infetti guariscano, è come se fossimo tutti immuni

Immunità di gregge procurata

Se introduciamo la vaccinazione in una popolazione durante la pandemia, consentiamo ad alcuni suscettibili di diventare direttamente ricoverati, senza passare per lo stato infetto



Ricordando $\text{nuovi infetti} = (b s(t) - k) i(t)$ possiamo affermare che la trasmissione di un virus termina quando la quota dei suscettibili è \leq di $1/R_0$. Nel caso del coronavirus $R_0=3$ e quindi la fine della pandemia viene raggiunta quando il 67% della popolazione è immune.

Ma se usiamo un vaccino efficace al 91%, allora l'immunità di gregge si raggiunge quando è stato vaccinato il 73%



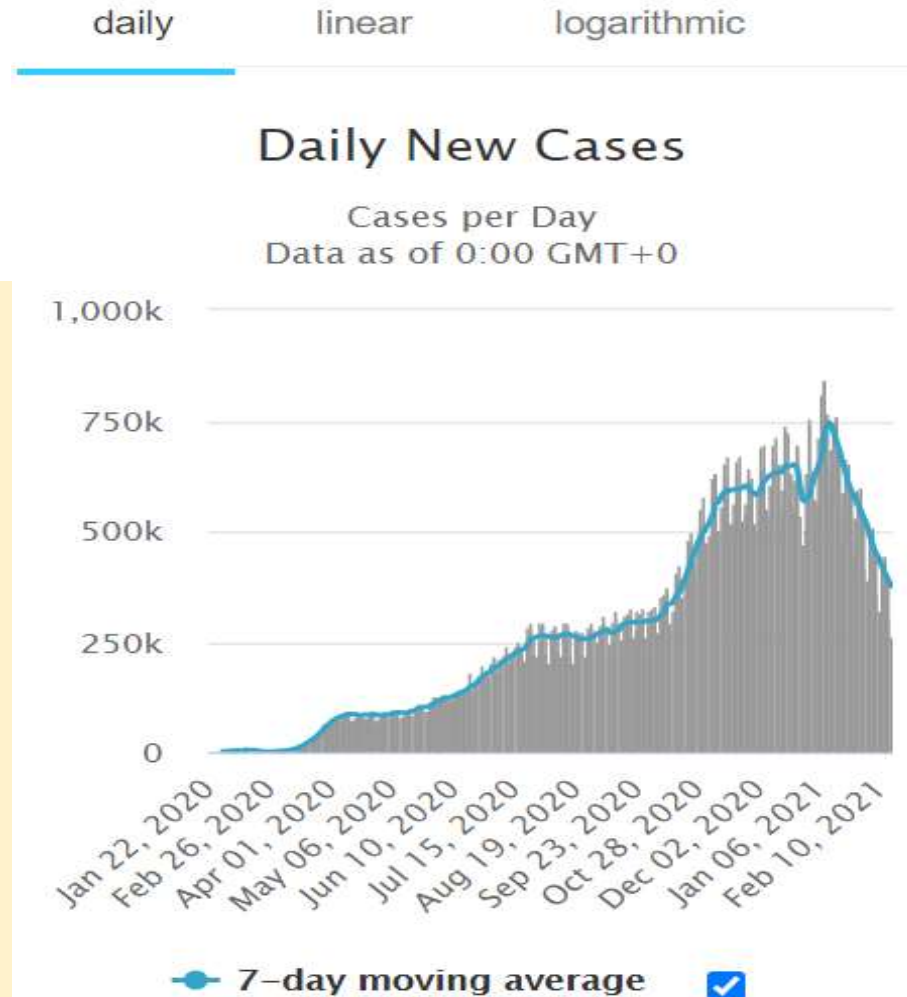
IL PROBLEMA DEL CASO REALE

Il modello SIR presuppone che i parametri **b e k** siano costanti nel tempo, che l'intera popolazione N soggetta alla pandemia sia un «ensemble» uniformemente distribuito su un dominio bidimensionale, con densità costante dei tre gruppi di individui e che nessun individuo si muova mai oltre il suo intorno locale

Ma il caso reale... è molto più complesso...

Andamento dei casi di pandemia nel mondo

Si noti la presenza di più picchi relativi ed una spiccata elevata oscillazione giornaliera, dovuta a fattori non pandemici ma organizzativi, compensata dalla media mobile a 7 g



Andamento degli infetti $I(t)$ testati positivi nella popolazione mondiale del Covid-19, in un anno per un totale di 110M di infetti su una popolazione di 7.700M di individui.

Un modello matematico realistico

I due parametri caratteristici b e k non sono costanti nel tempo, anche se questa ipotesi può essere non vera per k , perché k è tipico del virus che causa la pandemia, mentre il parametro b è sicuramente variabile nel tempo, ovvero $b=b(t)$, perché dipende dal comportamento sociale. Per descrivere meglio il fattore sociale che influenza l'andamento di una pandemia, poniamo

$$b(t) = \beta(t)\rho(t)$$

dove β è la probabilità di trasmettere il contagio per contatto aereo e ρ il numero medio di contatti giornalieri di un individuo infetto

$$\text{Nuovi infetti} = (\beta(t)\rho(t) s(t) - k) i(t)$$

Il fattore di previsione $R(t)$

Se facciamo il rapporto percentuale ogni giorno tra i nuovi casi e quelli totali e lo dividiamo per lo stesso rapporto del giorno precedente praticamente abbiamo una quantità che si avvicina a $R(t)$.

Ai fini pratici è sufficiente calcolare il valore percentuale della differenza della media mobile di 7 conteggi degli infetti, misurati in 7 giorni consecutivi per determinare $R(t)$

se $R(t) > 0$, il gruppo degli infetti aumenta, perché $\beta(t)\rho(t) > k$

se $R(t) < 0$, il gruppo degli infetti diminuisce, perché $\beta(t)\rho(t) < k$

Quando $\beta(t)\rho(t) = k$, allora $R(t)$ è uguale ad 1 e la curva dei contagi $I(t)$ raggiunge un punto critico (massimo/minimo), in quanto la differenza dei contagi giornalieri si annulla.

La previsione del picco

Nel caso $R(t)$ con media mobile su 7 giorni diminuisca per un certo numero di giorni, possiamo ipotizzare che questo trend continui nel futuro e che la variazione percentuale relativa dei contagi $r(t)$ diminuisca costantemente ogni giorno, di una quantità a , per stimare fra quanti giorni si ha il picco se il trend continuasse in questo modo, basta semplicemente fare la divisione $r(t)/a=n$.

Un poco più complesso è stimare il numero massimo di infetti con questo metodo veloce ma non è così difficile

Prevedere il carico sul sistema sanitario

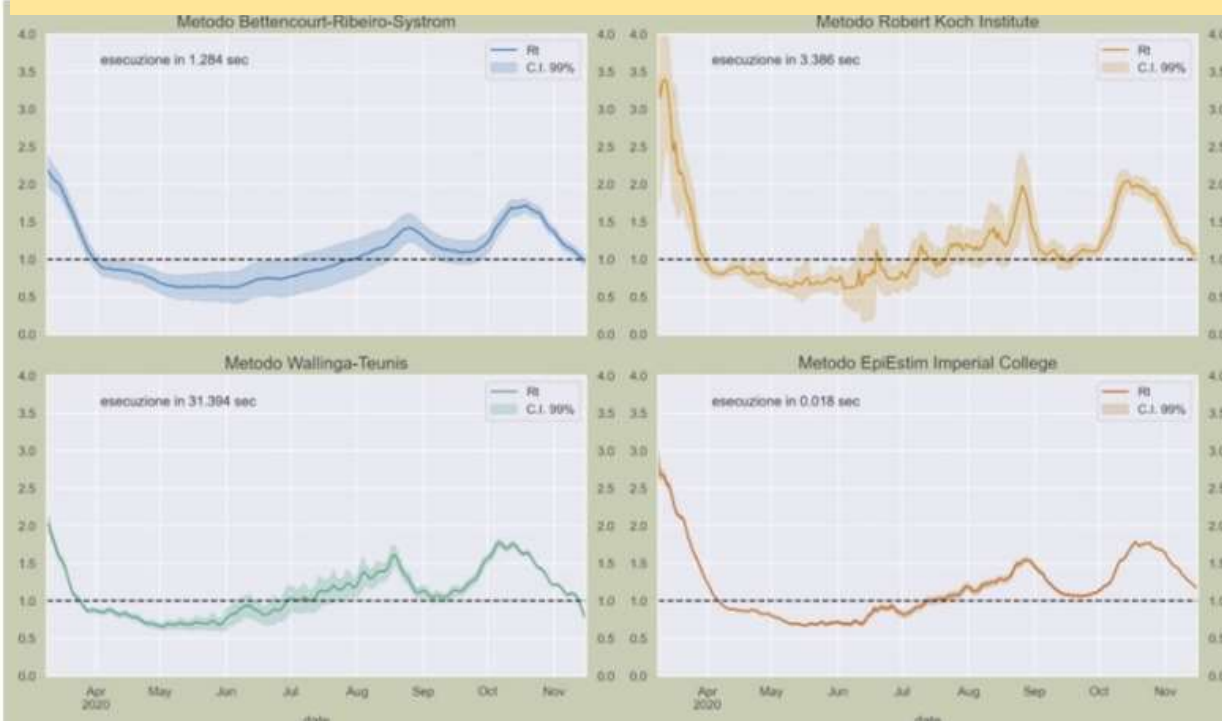
Nell'ipotesi che siamo in presenza in un trend negativo costante del tasso di variazione giornaliera meglio calcolare **la** media mobile degli ultimi sette giorni, per l'andamento tipico settimanale della rilevazione degli infetti, è possibile calcolare anche quale sarà il numero giornaliero di positivi al picco atteso e noto k anche quale sarà il numero totale infetti

Se sappiamo che una frazione f degli infetti viene ricoverata in strutture sanitarie Covid-19 per il trattamento, possiamo determinare il numero dei posti letto necessari ($I_{tot} \times f$) e decidere per tempo come intervenire

Chiaramente la previsione per estrapolazione sul trend del tasso di variazione giornaliera dei contagi, deve essere periodicamente rivista per adattare il modello previsionale al tasso corrente $R(t)$, che dipende dal comportamento sociale, come distanziamento, mascherine e lockdown.

Il tasso ufficiale $R(t)$

Il calcolo di $R(t)$ su una modellizzazione del contagio è basato su distribuzioni statistiche che descrivono la probabilità di avere uno o più contatti e la durata del periodo di contagio, incluso quello latente.



Come si può vedere nella figura a sinistra, per calcolare $R(t)$ esistono quattro metodi applicati al periodo dell'epidemia di Covid in Italia, che danno risultati che mostrano differenze anche significative l'uno dall'altro oltre ad avere diverse bande di incertezza. Questo rappresenta un problema quando il valore di $R(t)$ viene usato come parametro soglia per determinare le restrizioni che caratterizzano il colore di una regione. L'algoritmo che viene usato per implementare questi metodi è piuttosto complicato e fornisce sempre un $R(t)$ in ritardo rispetto al periodo esaminato. (cfr Roberto Battiston)

La complessità del caso reale

Per spiegare perché l'andamento del caso reale è diverso da quello atteso dal modello SIR dobbiamo prendere in considerazione fattori trascurati dal modello, il principale dei quali è che gli individui non vivono tutti insieme come un gregge ma vivono raggruppati in cluster mutuamente separati



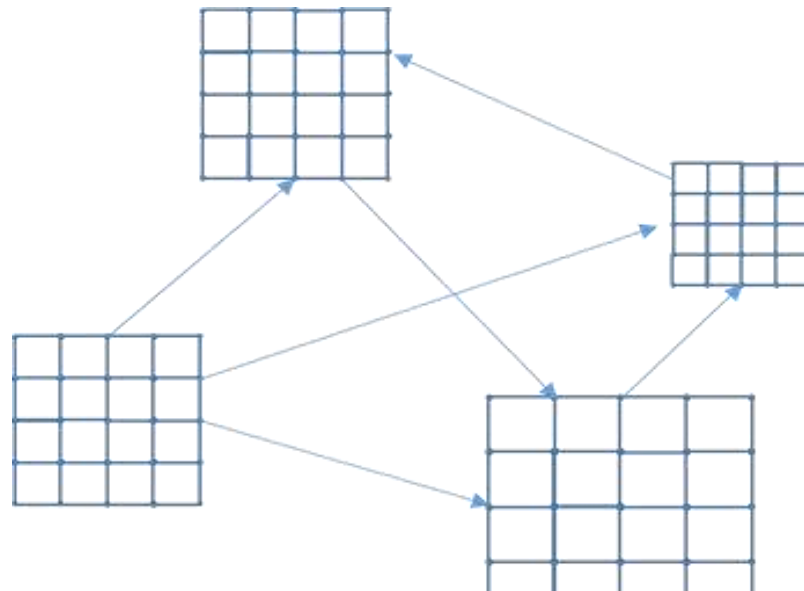
ed inoltre una grande percentuale di essi si sposta continuamente tra i vari cluster

Il modello multi-cluster

La popolazione reale non vive tutta ammassata in una unica, ma in cluster separati

I suscettibili non vivono mescolati «uniformemente» con gli infetti, ma esistono sacche omogenee per i vari gruppi

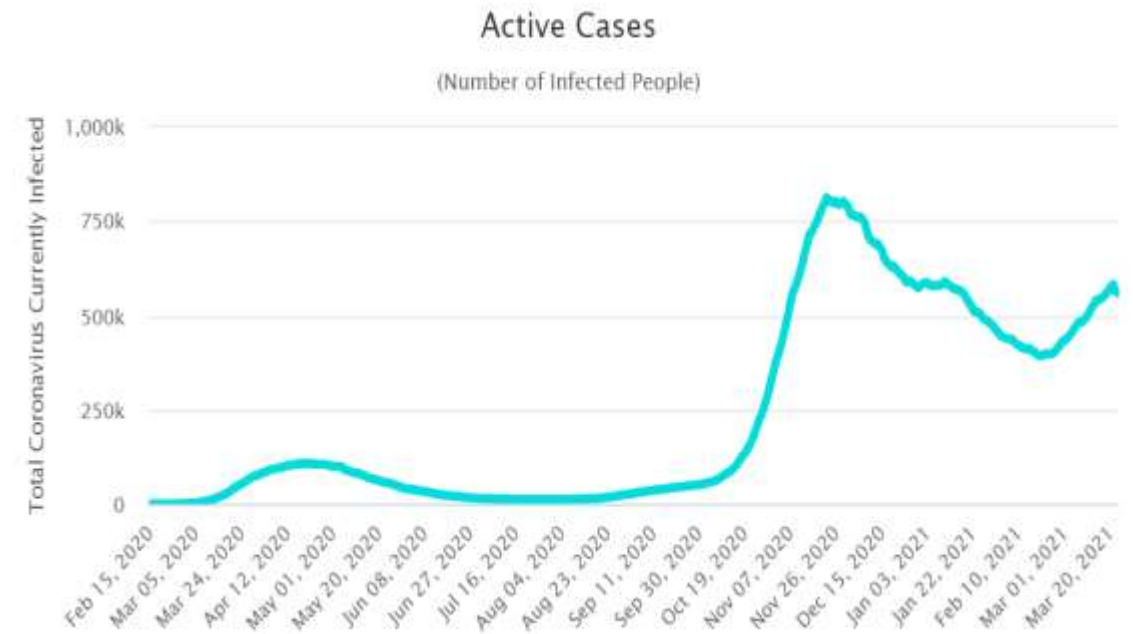
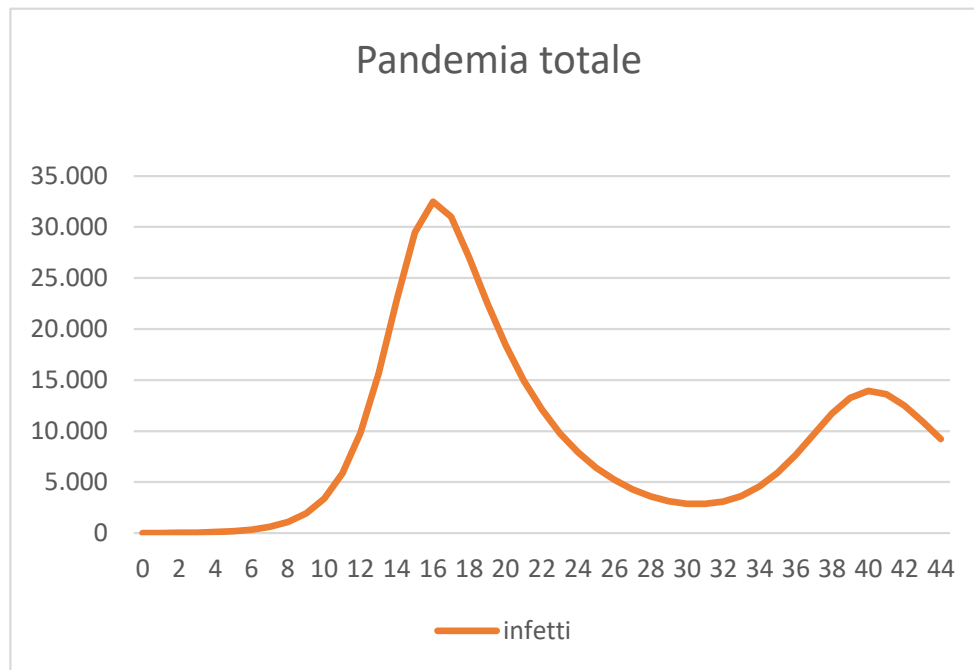
I cluster sono collegati da «ponti» formati da individui che si spostano tra i vari cluster in tempi asicroni



In ogni cluster si può applicare il modello dinamico SIR, la cui risultante è la curva tipica a molti picchi relativi, intervallati da periodi di minimo contagio.

Il multi-cluster

Nella figura seguente c'è l'andamento totale della pandemia in una popolazione divisa in tre cluster, dove un infetto ha diffuso nel cluster 2 dopo 15 giorni e nel cluster 3 dopo 20 giorni dall'inizio della pandemia scoppiata nel cluster 1



Come si può osservare in una popolazione multi-cluster appaiono le seconde ondate.

Come controllare $R(t)$

Mascherine,
sanificazione, social
distancing

Restrizioni, lockdown,
coprifuoco

$$R(t) = \beta(t) \rho(t)$$

$k(t)$

quarantena

Purtroppo controllare il comportamento contemporaneo di milioni di persone è complicato

Il tasso di variazione giornaliera (λ)

$$I(t) = e^{\lambda-1} I(t-1)$$

$$I(t)/I(t-1) = e^{\lambda-1}$$

$$\ln(I(t)/I(t-1)) = \lambda - 1$$

$$\ln(I(t)) - \ln(I(t-1)) + 1 = \lambda$$

Il tasso di incidenza giornaliera è correlato con $R(t)$ ma ha il vantaggio di essere tempestivo

CONCLUSIONI

- ❑ Il modello previsionale di una pandemia reale non funziona con l'ipotesi della popolazione come gregge
- ❑ L'andamento reale a molti picchi è sintomo che la popolazione vive in cluster a relativa bassa porosità
- ❑ L'indice $R(t)$ ha un'ampia incertezza statistica ed è sempre in ritardo rispetto al fenomeno che vuole descrivere
- ❑ La differenza relativa del tasso di incidenza è un indice tempestivo dell'evolversi della pandemia
- ❑ L'immunità di gregge è un concetto astratto per definire quando il fattore R diventa >1
- ❑ Con un'immunità di gregge al 60% si infetta ancora il 34,5% della popolazione
- ❑ Il piano vaccini deve essere tempestivo per avere una riduzione significativa degli infetti

APPENDICE

Per gli studenti del V anno è possibile analizzare le formule analitiche che regolano il modello matematico nelle slide seguenti

Il sistema di equazioni del modello SIR

Sistema
SIR

$$\frac{ds}{dt} = -b s(t) i(t),$$

$$\frac{di}{dt} = b s(t) i(t) - k i(t),$$

$$\frac{dr}{dt} = k i(t),$$

Il tasso di variazione (negativa) del numero dei suscettibili è proporzionale al numero degli stessi, al numero degli infetti corrente e al numero medio di contagi giornalieri, efficaci nella trasmissione

Il tasso di variazione del numero degli infetti è proporzionale al tasso di diminuzione dei suscettibili meno il tasso di variazione in aumento dei ricoverati

Il tasso di variazione dei ricoverati è proporzionale al tasso di guarigione giornaliera degli infetti

Il sistema è un sistema parametrico non lineare, per la mutua interazione tra le variabili dipendenti $s(t)$ ed $i(t)$. La variabile indipendente t è una variabile discreta, che misura il tempo in giorni.

Il metodo numerico di Eulero

Il metodo di Eulero permette di risolvere una equazione differenziale $dy(t)/dt$, nei punti discreti $y(0), y(1*\delta_t), \dots, y(n*\delta_t), \dots$ mediante la formula ricorsiva

$$y(n) = y(n-1) + \text{pendenza}_y(n-1) * \delta_t$$

E' quindi possibile usare un foglio di calcolo elettronico per descrivere il modello numerico SIR, simularne il funzionamento al variare dei parametri **b e k**

$$\begin{aligned} \frac{ds}{dt} = -b s(t) i(t), & \quad \longrightarrow \quad s_n = s_{n-1} - b s_{n-1} i_{n-1} \Delta t, \\ \frac{di}{dt} = b s(t) i(t) - k i(t), & \quad \longrightarrow \quad i_n = i_{n-1} + (b s_{n-1} i_{n-1} - k i_{n-1}) \Delta t, \\ \frac{dr}{dt} = k i(t), & \quad \longrightarrow \quad r_n = r_{n-1} + k i_{n-1} \Delta t. \end{aligned}$$

Dove occorre specificare i valori di: **b, k, s(0), i(0), r(0)** noto **N**

Immunità di gregge naturale

Se dividiamo la seconda equazione del sistema per la prima otteniamo, $c=b/k$:

$$\frac{di}{ds} = -1 + \frac{1}{cs}$$

Che una volta risolta fornisce la quantità seguente indipendente dal tempo t , (uguale a 1 per $t=0$)

$$i + s - \frac{1}{c} \ln s$$

Quando la pandemia è decisamente finita, abbiamo $i(\text{infinito})=0$ e $s(\text{infinito})$ uguale agli individui che non si infettano più, ottenendo la relazione

$$c = \frac{\ln s_{\infty}}{s_{\infty} - 1}$$

La misura della contagiosità

Ricordando che **b** indica quanti contagi sono generati da un infetto in un giorno e **k** quanti giorni si è contagiosi per diffondere quel virus, possiamo definire un parametro cumulativo

$$c = b/k$$

c indica quanto contagi sono generati da un infetto e quindi dà una misura della contagiosità di un virus, che dipende sia dal virus (**k**) che dal comportamento della popolazione (**b**) ed è indicato con **R0**.

Per poter determinare quanto vale **c** possiamo usare il calcolo analitico della funzione

$$di = (b s(t) i(t) - k i(t)) dt$$

che è la seconda funzione del modello SIR, ed indica quanti nuovi infetti ci sono nell'unità di tempo.

Tale funzione ha un massimo quando $b s(t) i(t) - k i(t) = 0$ ovvero $b s(t) - k = 0$ il che implica anche trovare $s(t_{max}) = k/b = 1/c$, da cui dedurre **c**

La previsione del picco

Nel caso $R(t)$ con media mobile su 7 giorni diminuisca per un certo numero di giorni, possiamo ipotizzare che questo trend continui nel futuro e che la variazione percentuale $r(t)$ uguale a $dl(t)/l(t)$ diminuisca linearmente con il tempo, tramite il fattore a , possiamo porre la condizione della 1) per trovare il valore di a .

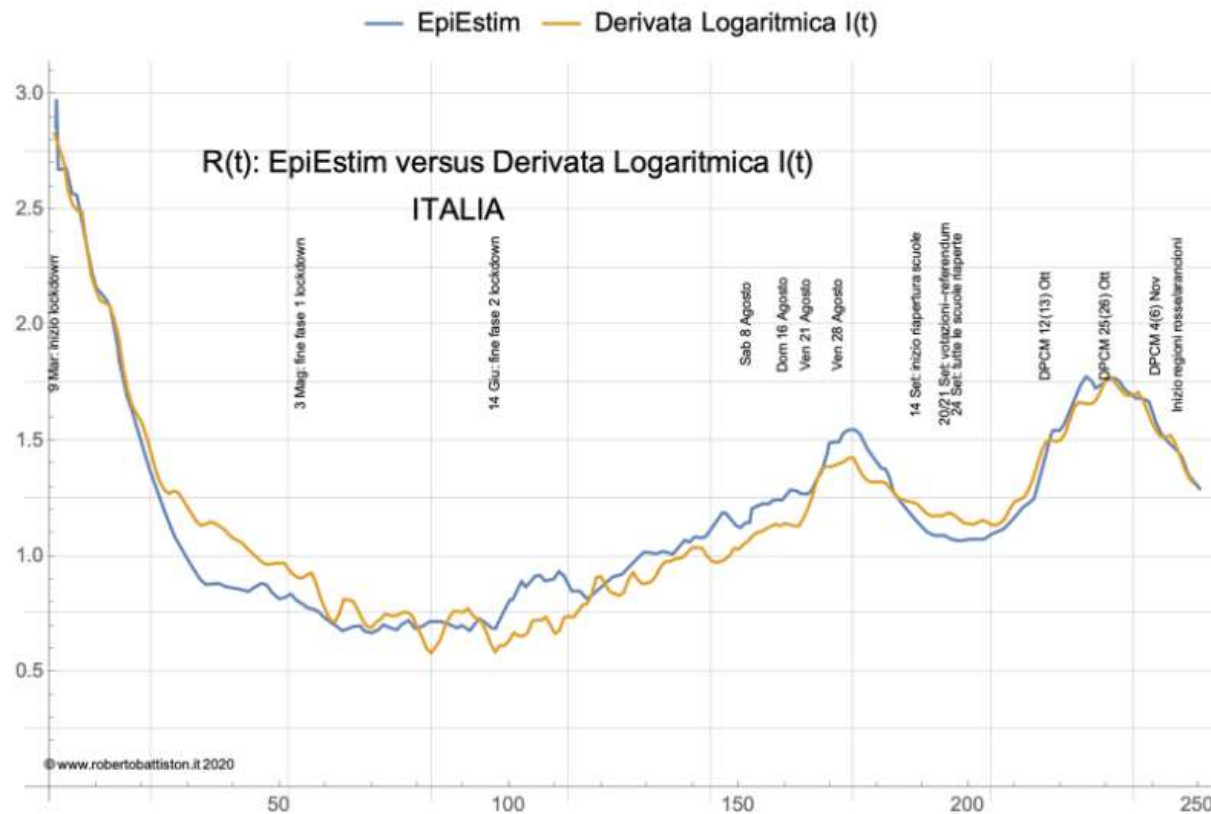
1)

$$\int_{r_{in}}^{r_{fi}} \frac{dy}{y} = -a \int_0^n dt$$

La soluzione è $\ln(r_{fi}) - \ln(r_{in}) = -a n$ e quindi determiniamo $a = \ln(r_{in}) - \ln(r_{fi}) / n$, dove a è il fattore caratteristico dell'ondata corrente della pandemia sotto analisi. Noto a , se vogliamo prevedere quando si raggiungerà il picco è sufficiente, porre $\ln(r_{in}) = a n_{max}$, e determinare n_{max} in giorni calcolando $n_{max} = \ln(r_{in}) / a$.

R(t) logaritmico verso EpiEstim

Il metodo speditivo logaritmico non dà risultati molto distanti da EpiEstim usato da ISS ma permette di valutare R(t) ogni singolo giorno appena sono disponibili i dati sui nuovi positivi



Tratto dal sito di Roberto Battiston.