La curva dei contagiati da COVID-19: la ricerca del punto di svolta.

scritto da Vito M. R. Muggeo | 17 Marzo 2020



L'intero paese sta con il fiato in sospeso in attesa di segnali del rallentamento dell'epidemia di coronavirus. Vito Muggeo e Mariano Porcu propongono un metodo statistico per individuare i momenti nei quali la curva di crescita dei casi di infezione rivela cambi di passo che preludono al rallentamento del processo di diffusione del virus.

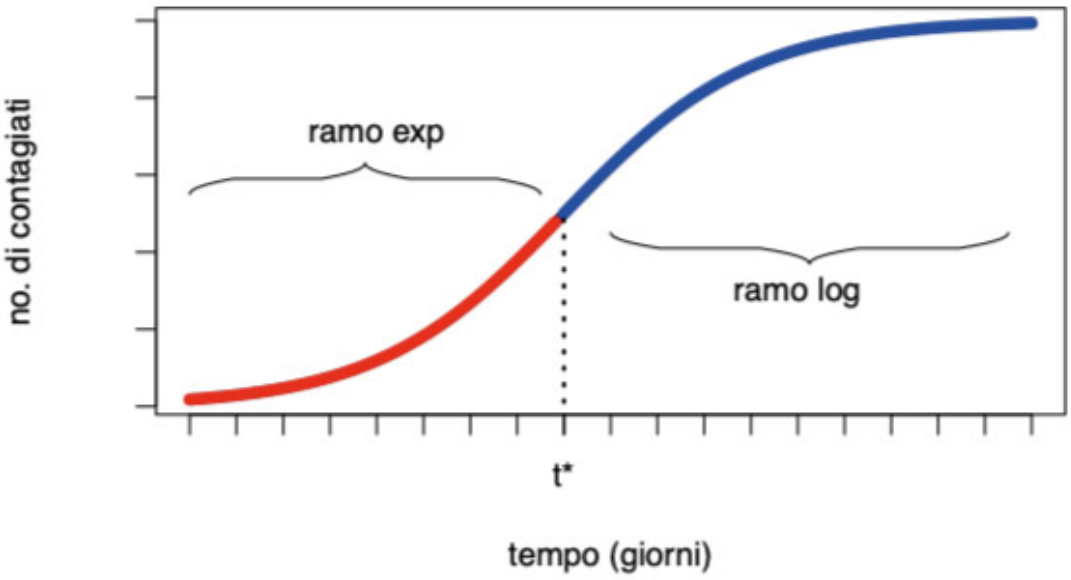
In queste giornate dell'epidemia da COVID-19 si sente tanto discutere della "curva dei contagiati", ovvero di come il numero complessivo dei soggetti positivi, indipendentemente dal loro stato attuale (guariti, ancora ammalati o, purtroppo morti), evolve di giorno in giorno o, più generalmente, rispetto al tempo. Tra le statistiche che vengono comunicate ufficialmente dalla Protezione Civile ad ogni fine giornata, il numero dei conteggi dei positivi e la corrispondente curva dei contagiati è lo strumento che viene utilizzato dagli epidemiologi e addetti ai lavori per monitorare l'andamento dell'epidemia. La curva dei contagiati, in pratica, svolge la stessa funzione di uno strumento nel cruscotto di una macchina, ad esempio, un aereo: il pilota osserva lo strumento che *indica* come sta andando il suo apparecchio. Ovviamente, un aereo è una macchina complessa, con tanti strumenti nel *cockpit* della cabina di pilotaggio. Anche una malattia contagiosa è un sistema complesso e la curva dei contagiati è uno degli strumenti nel cruscotto a disposizione dei nostri medici e epidemiologi per seguirne la diffusione.

La logistica non è solo quella del business

Nella letteratura in ambito statistico ed epidemiologico, la forma della

curva dei contagiati è ben nota, ed è quella riportata nella Figura 1.

Figura 1



È una curva a forma di S allungata, nota in letteratura come curva logistica. È formata da due rami: uno esponenziale (exp) e l'altro logaritmico (log) uniti in corrispondenza di un punto t* del quale parleremo più avanti. La curva logistica è largamente utilizzata per descrivere la crescita di popolazioni, sia in generale[1] sia, come in questo caso, dei contagiati da COVID-19. All'inizio della diffusione dell'epidemia l'andamento è di tipo esponenziale (exp): il numero dei contagiati cresce con incrementi giornalieri progressivamente crescenti e quasi sembra esplodere. Superato un certo istante t*, che rappresenta il punto di flesso della curva, il numero dei contagiati continua ad aumentare, ma più lentamente, vale a dire con incrementi che diminuiscono con il passare dei giorni. Alla fine del periodo di osservazione il numero dei contagiati non crescerà più e la diffusione della malattia potrà considerarsi in via di esaurimento. Ci sono essenzialmente due ragioni che portano al passaggio dal ramo exp a quello log, ed infine al non avere nuovi casi di contagiati:

(1) la popolazione di riferimento si satura progressivamente fino a quando tutti sono contagiati, oppure, caso più realistico,

(2) vengono messe in atto misure di contenimento finalizzate alla riduzione dei contagi.

Osservare quando lo strumento nel nostro cruscotto, vale a dire la curva di crescita, si sta avvicinando al punto t* ha un'importanza fondamentale:

significa capire che la corsa esponenziale e apparentemente inarrestabile del contagio sta rallentando e si sta raggiungendo un punto di svolta a partire dal quale le nostre armi contro la diffusione della malattia iniziano a vincere. L'istante t*, in pratica, per il suo valore di indicatore di cambio della velocità di diffusione (la crescita della malattia passa da exp a log), rappresenta la quantità di fondamentale interesse a cui tutti speriamo si arrivi prima possibile.

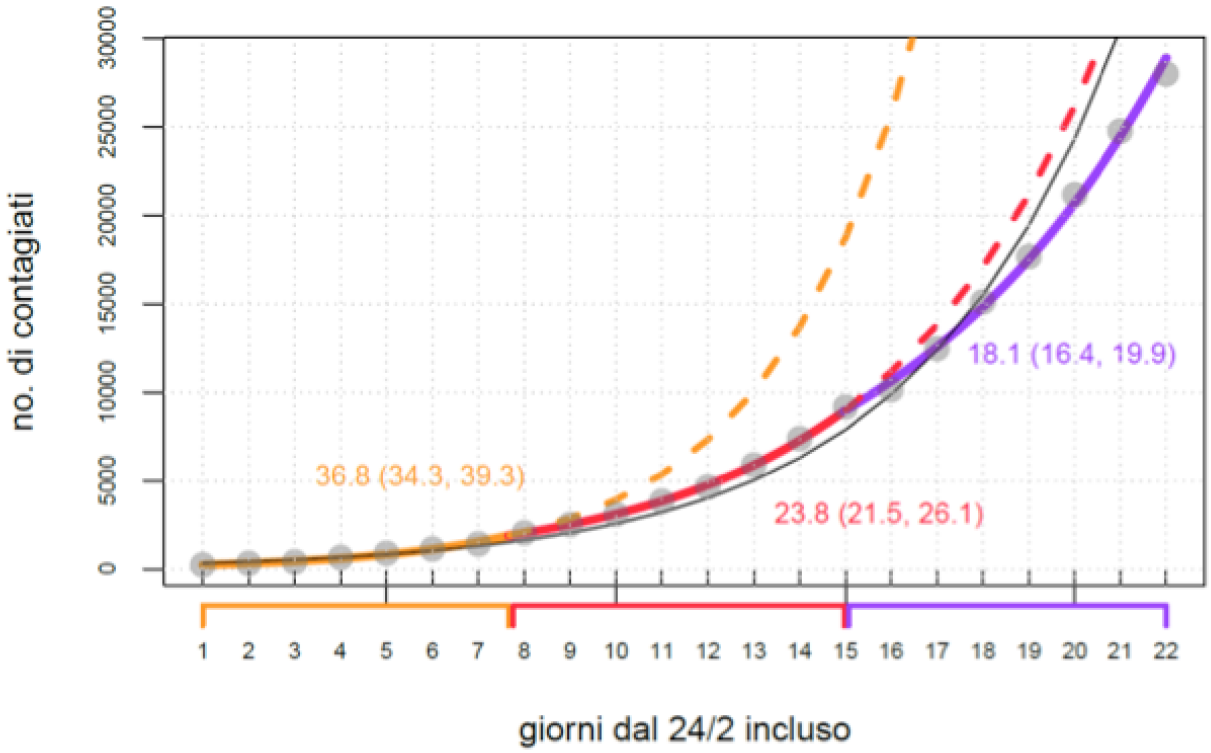
t* … cerco

Purtroppo, conoscere questo istante t* è impresa ardua e probabilmente impossibile, in quanto dipende dall'efficacia delle misure di contenimento (che, nel caso del COVID-19 dipendono essenzialmente dalla collaborazione della popolazione), nonché dal numero degli individui che si sono contagiati prima della messa in atto di tali misure e non ultimo, anche dal periodo di incubazione della malattia, che per il COVID-19 risulta essere molto variabile, tra 2 e 12 giorni con un tempo medio di 5,5 giorni[2]. Tuttavia, qualche segnale di avvicinamento all'istante t* può essere captato anche studiando l'andamento iniziale exp della curva dei contagiati: ovvero, pur trovandoci nella parte in cui il trend è "esplosivo", è verosimile pensare che una parte della popolazione sia stata più reattiva alle misure di contenimento, magari le abbia anticipate per propria iniziativa, portando ad un leggero rallentamento del trend. In pratica, questo significherebbe che la crescita esponenziale non sta avvenendo con un unico tasso di incremento ma con tassi variabili decrescenti nel tempo.

Modelli statistici semplici e (utilmente) complicati

Analizziamo il numero complessivo giornaliero dei contagiati da COVID-19 in Italia a partire dal 24 febbraio fino al 16 marzo. Un andamento esponenziale può essere ottenuto attraverso un'unica equazione di regressione che esprime (o modella) il numero di contagiati in funzione del tempo, contato in giorni. Sembra però più adatto al caso italiano un modello di regressione *segmented*[3], in cui l'andamento del numero dei contagiati (in logaritmo) viene approssimato con segmenti uniti, ma di pendenze diverse. La Figura 2 mostra i risultati ottenuti, insieme ai tassi di crescita esponenziali dei tre sotto-periodi identificati.

Figura 2 – Numero dei contagiati dal 24/02/2020



Guida alla lettura. Il tasso di crescita del modello esponenziale (segmented) che meglio approssima l'evoluzione del numero dei contagiati è venuto calando nel tempo: da un alto valore iniziale (36.8) si è attualmente (16 marzo 2020) ridotto a 18.1. In corrispondenza, il tempo di raddoppio si è allungato: se prima bastavano circa 2 giorni perché il numero dei contagiati raddoppiasse, ora ne occorrono circa 4. Nella figura sono anche mostrati i valori dell'intervallo di confidenza al 95%

Ogni punto grigio è il conteggio effettivo giornaliero: come la teoria suggerisce, l'andamento di fondo è indubbiamente esponenziale con incrementi giornalieri crescenti. Tuttavia un modello con un unico tasso di crescita, quello medio finora osservato, sembra fornire un adattamento non soddisfacente (linea nera sottile nel grafico), con differenze tra valori osservati e stimati abbastanza rilevanti soprattutto in corrispondenza delle ultime osservazioni. La modellazione *segmented* invece sembra adattarsi meglio ai dati[4]. Il modello stesso identifica (o meglio, *stima*) due istanti temporali (tecnicamente "breakpoints") in corrispondenza dei quali il tasso di crescita esponenziale ha cambiato forza, diminuendo. Il tasso del primo periodo, il più elevato, che ha caratterizzato l'andamento (arancione nel grafico) fino all'ottavo giorno (2 marzo, approssimativamente), è diminuito (rosso) fino al 15° giorno (9 marzo) quando si è verificata un ulteriore riduzione portando ad un andamento esponenziale con un tasso di incremento ancora più ridotto (viola)[5]. Il grafico evidenzia attraverso le linee tratteggiate anche l'andamento che si sarebbe avuto se non si fossero registrate queste riduzioni. In particolare, la linea arancione tratteggiata potrebbe essere interpretata come l'andamento "naturale" della diffusione del COVID-19, ovvero quello che verosimilmente avremmo osservato in mancanza delle prime precauzioni e restrizioni: ad es., alla data dell'8 marzo (giorno

14) circa 14.000 contagiati.

Cosa se ne trae

Il messaggio conclusivo di questa nota è che, grazie a un'analisi *segmented* come quella qui brevemente tratteggiata, potrebbe essere importante riuscire a cogliere anche i più piccoli segnali di cambiamento nel trend esponenziale. Ovviamente per popolazioni molto grandi (ad esempio la Cina) i cambiamenti nella parte esponenziale della crescita potrebbero essere impercettibili e non rilevanti da un punto di vista pratico. Per popolazioni di medie dimensioni, come l'Italia, invece, i cambi nel ramo esponenziale della curva, anche se contenuti, potrebbero avere importanti effetti sulle politiche sanitarie di gestione delle cure per i contagiati in più gravi condizioni.

Gli aggiornamenti della figura 2 di questo articolo sono disponibili qui:

Note

[1] Il modello di crescita logistica è anche chiamato modello di Verhulst. Un'introduzione al modello esponenziale (e alla crescita logistica) può essere trovato qui.

[2]Stephen A. Lauer, MS, PhD; Kyra H. Grantz, BA; Qifang Bi, MHS; Forrest K. Jones, MPH; Qulu Zheng, MHS; Hannah R. Meredith, PhD; Andrew S. Azman, PhD; Nicholas G. Reich, PhD; Justin Lessler, PhD The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application

[3] V.M.R. Muggeo (2003), Estimating regression models with unknown break-points, *Statistics in Medicine*, 22, 19.

[4] Il modello segmented con 2 punti di svolta (breakpoints) ha mostrato un adattamento migliore anche di altri modelli basati su funzioni logistiche generalizzate (cioè caratterizzate da più parametri).

[5] In corrispondenza della riduzione dei tassi si è avuta anche un aumento dei *tempi di raddoppio*, (ovvero il numero di giorni necessari per duplicare il numero dei contagiati) che nei 3 sotto-periodi risultano essere 2,2, 3,2 e 4,1 giorni.