

Alcuni risultati dell'analisi dei dati epidemiologici del Coronavirus in Italia

G. Sebastiani ¹

I dati qui analizzati sono stati scaricati dal sito <https://github.com/pcm-dpc/COVID-19>. Tra i vari modelli matematici a disposizione per descrivere il fenomeno della diffusione dell'epidemia del coronavirus, uno dei più semplici coinvolge alcuni “compartimenti”. Il primo compartimento \mathcal{S} contiene gli individui che non presentano il virus all'interno del loro corpo e che risultano quindi “suscettibili” di essere infettati. Da questo compartimento ciascun individuo può spostarsi nel secondo compartimento \mathcal{I} dove ci sono invece gli individui che sono stati infettati. Da qui ciascun individuo può andare in un terzo compartimento \mathcal{M} che contiene i soggetti che hanno sviluppato i sintomi della malattia, diagnosticata di conseguenza tramite test e che siano stati isolati, a casa, in ospedale o in terapia intensiva. Da questo compartimento, un individuo passerà nel quarto ed ultimo compartimento dove si trovano gli individui guariti e quelli morti. Assumiamo qui che i guariti non contengano più il virus, o che comunque non possano trasmetterlo ad altri individui infettandoli. È comunque stato riportato che in Cina circa il 14% dei pazienti guariti dall'infezione da coronavirus presentano il virus nelle feci dopo due settimane dalla dimissione. Osserviamo che non tutti gli individui infetti passano nel compartimento \mathcal{M} . Questo può accadere sia per una mancata o erronea diagnosi e/o isolamento. Più preoccupante è il caso in cui una frazione non trascurabile degli infetti non sviluppino i sintomi della malattia, pur potendo trasmetterla, cioè siano “portatori sani”. Questo sembra essere il caso per il coronavirus. Purtroppo, da circa la fine di Febbraio non vengono più effettuati test diagnostici a campione sugli asintomatici che permetterebbero una stima della frazione dei portatori sani.

Consideriamo le quattro funzioni $S(t)$, $I(t)$, $M(t)$ e $R(t)$ che descrivono il numero di individui presenti al tempo t in ciascuno dei quattro compartimenti. A questo punto possiamo ragionare in termini deterministici considerando il flusso di spostamento da \mathcal{S} ad \mathcal{I} , quello da \mathcal{I} ad \mathcal{M} e quello da \mathcal{M} ad \mathcal{R} . Il primo flusso può essere assunto proporzionale sia al numero di individui suscettibili $S(t)$ che al numero di infetti $I(t)$. Gli altri due flussi possono invece essere assunti proporzionali ad $I(t)$ e ad $M(t)$ rispettivamente. Analogamente, possiamo ragionare in modo probabilistico rimpiazzando i flussi con delle probabilità. Nel caso deterministico, le tre funzioni possono essere determinate risolvendo numericamente un sistema di equazioni differenziali ordinario. Nel caso stocastico possiamo usare invece l'algoritmo di simulazione di Gillespie.

Ci siamo finora focalizzati sul numero $C(t) = M(t) + R(t)$ di persone che sono state

¹Istituto per le Applicazioni del Calcolo “Mauro Picone”, Consiglio Nazionale delle Ricerche, Rome, Italy, Mathematics department “Guido Castelnuovo”, “Sapienza University of Rome”, Italy, Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Roma 1, Rome, Italy, and Department of Mathematics and Statistics, University of Tromsø, Norway

contagiate ed osservate fino al tempo t . Abbiamo optato per questa scelta sperando che gli errori da cui sono affette le misure delle singole variabili, e.g. il numero dei morti, di cui $C(t)$ è la somma, si compensino. La funzione $C(t)$ è per definizione crescente nel tempo, o meglio non decrescente. Inoltre, dalle soluzioni ottenute numericamente, osserviamo un aumento progressivo del tasso di crescita di $C(t)$ fino ad un massimo oltre il quale tale tasso diminuisce, mentre $C(t)$ continua a crescere ma sempre più lentamente e dopo un tempo sufficientemente lungo non ha in pratica più variazioni. Una funzione che ha questo andamento è la funzione “logistica” $y_0/(1 + \exp(-(t - t_0)/\tau))$, come mostrato in Figura 1. Dei tre parametri y_0 , t_0 e τ contenuti in questo modello, il parametro positivo τ influenza la velocità con cui viene raggiunto il valore limite: diminuendo il suo valore, aumenta tale velocità.

In Figura 2 si osserva la sequenza temporale della frazione dei contagiati osservati in Italia rispetto al numero totale di abitanti, assieme ad un fit con due modelli matematici applicati in sequenza in ciascuno di due sotto-intervalli contigui in cui abbiamo diviso l’intervallo temporale. Il modello utilizzato in ciascun sotto-intervallo è di tipo geometrico $y_0 2^{t/\tau}$. Questa funzione, al contrario della logistica, cresce indefinitivamente e così fa anche il suo tasso di crescita. Il parametro positivo τ rappresenta il tempo di raddoppio del valore della funzione. Più basso è il valore di questo parametro, più rapidamente cresce nel tempo il valore della funzione. La scelta di usare due modelli in sequenza è stata motivata dal cambiamento del criterio di identificazione dei contagiati avvenuto alla fine del mese di Febbraio, che non coinvolge più l’uso di tamponi su soggetti asintomatici. Sfortunatamente, si intuisce facilmente che la frazione dei contagiati effettivi dopo il cambio di criterio risulta significativamente sottostimata.

Sulla base dei dati osservati fino ad oggi, a livello dell’intera Italia non c’è evidenza di riduzione del tasso di crescita dei contagiati osservati. Infatti l’adattamento del modello ai dati è equivalente a quello in cui la funzione geometrica nel secondo sotto-intervallo è sostituita dalla logistica. Questo potrebbe essere influenzato da una stima per difetto del numero di contagiati osservati fino al 10 Marzo in Lombardia, che contribuisce per circa il 60% al numero totale di contagiati osservati in Italia. Si spera che questo possa essere chiarito con l’analisi dei dati aggiornati a domani pomeriggio 12 Marzo alle 17. Anche per la Lombardia non si osserva evidenza di diminuzione del tasso di crescita della frazione dei contagiati osservati, come si può apprezzare in Figura 3.

Comunque, sulla base dei dati aggiornati all’11 Marzo alle 17, la sequenza della frazione dei contagiati osservati nelle altre 11 regioni per cui $C(t)$ supera il valore di 50 (escludendo le province del Trentino) mostra evidenza di diminuzione del tasso di crescita. Infatti il modello misto geometrico-logistico si adatta meglio ai dati di quello doppio geometrico. In Figura 4 possiamo osservare il risultato relativo al Piemonte.

I valori stimati del tempo dal 1 Marzo a partire dal quale non si avranno praticamente variazioni della frazione dei contagiati osservati sono riportati in Tabella 1. Questi valori sono delle stime per difetto dei veri valori del tempo di stabilizzazione. C’è grande incertezza su questi valori a causa del limitato intervallo temporale in cui il tasso di crescita diminuisce. Col tempo tale intervallo aumenterà e l’incertezza della stima diminuirà. Naturalmente questo vale se non ci saranno variazioni significative del comportamento della popolazione in relazione ai meccanismi principali alla base del fenomeno di diffusione del virus.

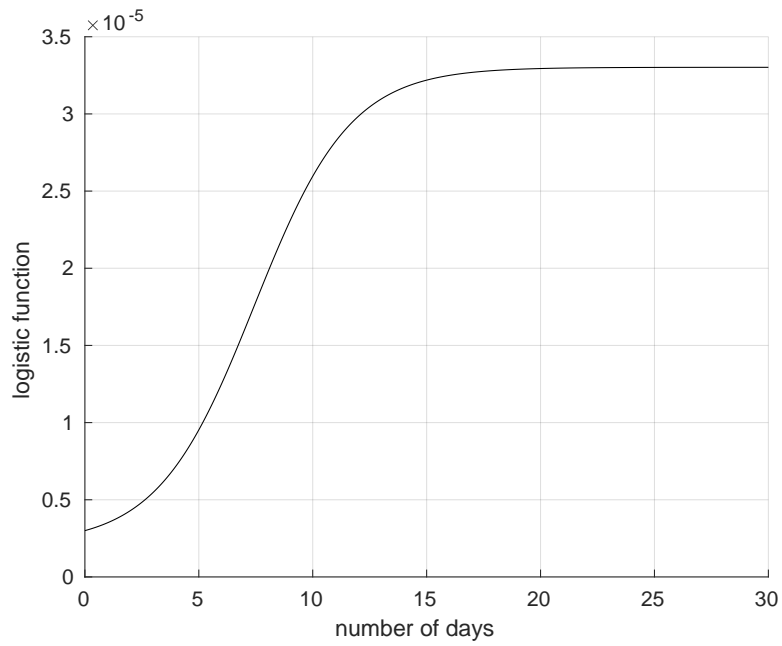


Figure 1: Esempio di andamento temporale di una variabile descritta da una funzione logistica.

Em.	Ven.	Piem.	Mar.	Tosc.	Lig.	Camp.	Laz.	Fri.	Sic.	Pug.
30	45	15	40	30	20	20	80	15	30	25

Table 1. Valori stimati del tempo di stabilizzazione (in giorni) della frazione di contagiati osservati per le 11 regioni considerate.

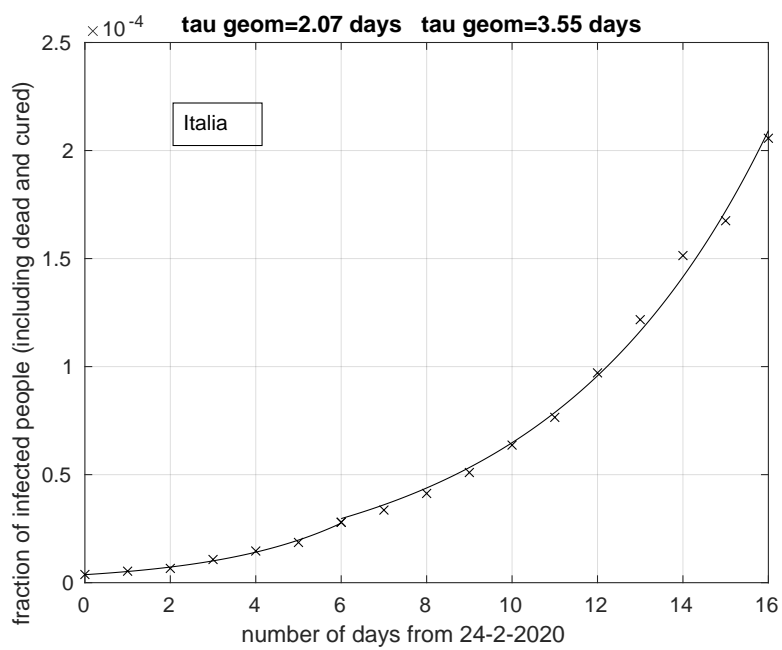


Figure 2: Sequenza della frazione del numero dei contagiati osservati in Italia rispetto alla popolazione nazionale. Il miglior fit con un modello doppio geometrico è sovrapposto ai dati.

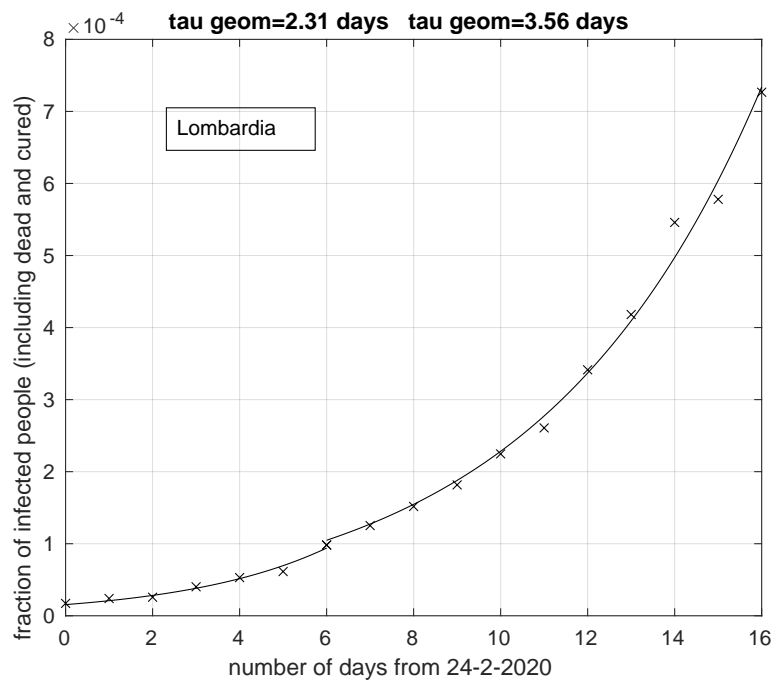


Figure 3: Sequenza della frazione dei contagiati osservati in Lombardia rispetto alla popolazione della regione. Il miglior fit con un modello doppio geometrico è sovrapposto ai dati.

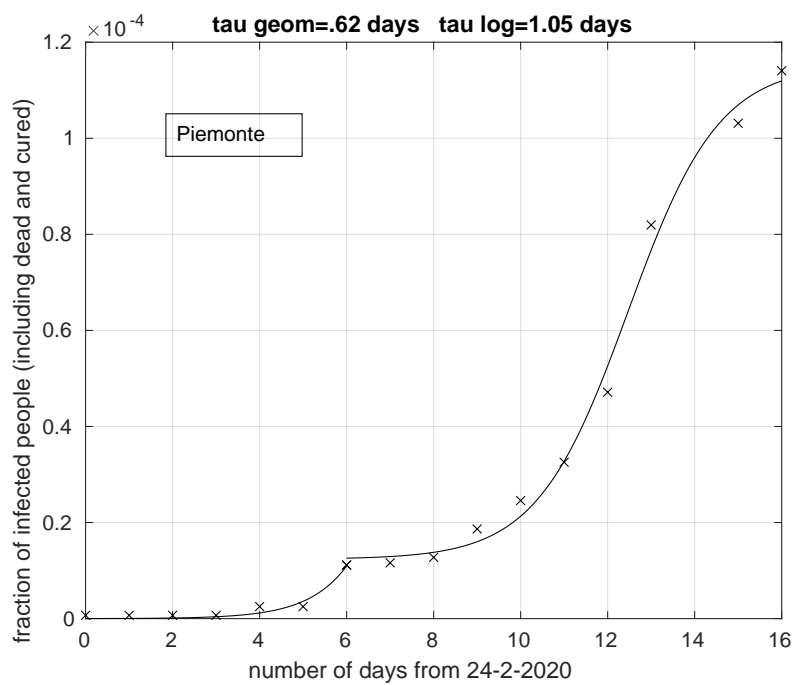


Figure 4: Sequenza della frazione dei contagiati osservati in Piemonte rispetto alla popolazione della regione. Il miglior fit con un modello misto geometrico-logistico è sovrapposto ai dati.