

Come stimare la fine dell'emergenza Covid-19 in Italia

by Emanuele¹ e Luigi² Vannucci

¹Università di Pisa ²Università di Firenze

1. Epidemie e curve logistiche

E' ben noto che molte epidemie ex post possono essere rappresentate con la funzione logistica o con somme di logistiche. Queste funzioni hanno come variabile indipendente il tempo, spesso con il giorno quale unità di misura, e come variabile dipendente il numero cumulato di contagiati o di guariti o di ospedalizzati o di decessi... alle varie date.

Ex post, a epidemia conclusa, avendo a disposizione i dati giornalieri è un banale problema l'interpolazione dei dati con la "miglior" funzione della famiglia di funzioni prese in considerazione: imponendo la minimizzazione della somma dei quadrati degli scarti tra i dati empirici e i valori letti sulla funzione interpolante si otterrebbero in modo univoco i valori da dare ai parametri per individuare la miglior funzione di interpolazione e da questi valori ricavare poi alcuni indicatori di sintesi caratterizzanti l'epidemia.

Se si assume di ricercare la migliore funzione interpolante nella famiglia delle funzioni logistiche allora si deve considerare la

$$f(t) = \frac{c}{1 + \exp(-a \cdot (t - b))}$$

in cui sono presenti tre parametri reali positivi a , b , c , con t il tempo, si supponga misurato in giorni, e con $f(t)$ il numero dei decessi (ma potrebbe essere quello dei contagiati...) fino all'epoca t .

Si osservi che si ha

$$f(t) = c - f(2b - t)$$

da cui si segue che il numero di decessi associati a due intervalli di ugual lunghezza e in simmetria con l'asse $t = b$ è lo stesso.

La derivata

$$f'(t) = \frac{ac \cdot \exp(-a \cdot (t - b))}{(1 + \exp(-a \cdot (t - b)))^2}$$

verifica ovviamente

$$f'(t) = c - f'(2b - t)$$

L'intensità di morte, se p è il numero totale dei membri della popolazione in cui si accende l'epidemia, è data da

$$\mu(t) = \frac{f'(t)}{p - f(t)} = \frac{\frac{ac \cdot \exp(-a \cdot (t - b))}{(1 + \exp(-a \cdot (t - b)))^2}}{p - \frac{c}{1 + \exp(-a \cdot (t - b))}}$$

Curiosamente, se l'epidemia sterminasse l'intera popolazione ovvero se $p = c$, si avrebbe

$$\mu(t) = \frac{a}{c} \cdot f(t)$$

ovvero intensità di morte sarebbe proporzionale alla stessa funzione logistica.

Tenuto conto di queste osservazioni, per questa famiglia di funzioni è molto semplice associare ai valori dei parametri stimati dai dati empirici gli indicatori che caratterizzano l'epidemia, poiché

- c è il numero totale dei decessi nell'intera durata dell'epidemia.
- b è il cosiddetto "picco" dell'epidemia, esso indica il giorno, rappresentato dall'intervallo temporale $[b, [b + 1]$, in cui si registra il massimo dei decessi. Indica anche l'epoca in cui il processo è a metà percorso, o emidurata, ($f(b) = \frac{c}{2}$). Se si misura il tempo in giorni in modo che in 0 sia $f(0) = \frac{c}{1 + \exp(ab)}$ "trascurabile" (ovvero < 1 se si contano i morti) allora $2b$ è interpretabile come l'intera durata in giorni dell'epidemia (ovviamente per quanto prima si è visto $c - f(2b) = f(0)$).
- a è l'indicatore dell'intensità con cui il processo procede al picco e nei giorni contigui al picco. Il differenziale $f'(b) \cdot 1 = \frac{ac}{4}$ indica con buona approssimazione il numero di morti nell'intervallo $[b, b + 1]$ e $\frac{a}{4}$ è dunque la quota di decessi avvenuti in quell'intervallo lungo un giorno. Per esempio se $a = 0.8$ nel giorno $[b, b + 1]$ si avrebbe in sostanza $\frac{1}{5}$ del numero totale dei decessi.

La funzione logistica si rivela efficace quando l'epidemia ha un unico focolaio che non si diffonde sul territorio, ma quando ci sono più focolai che si accendono in tempi diversi la situazione si complica e un modello più generale, pensando a m focolai distinti, consiste nella considerazione di funzioni ottenute dalla somma di m logistiche, ognuna caratterizzata dai suoi tre parametri a_h, b_h, c_h per $h = 1, 2, \dots, m$. Analiticamente si direbbe considerare per il conteggio del numero dei decessi fino all'epoca t la funzione

$$g(t) = \sum_{h=1}^m \frac{c_h}{1 + \exp(-a_h \cdot (t - b_h))}$$

e si avrebbero $3m$ parametri a disposizione per la ricerca della migliore interpolazione di dati aggregati, ma con difficoltà crescenti al crescere di m nella implementazione di algoritmi per la ricerca del minimo della funzione di penalizzazione, qualunque essa sia!

2. Emergenza Covid-19

L'emergenza Covid-19, che ci ha cambiato le abitudini e i modi di vivere in questo anno 2020, è sorta come epidemia nella città di Wuhan in Cina, ma è poi diventata una pandemia interessando le popolazioni dell'intero Pianeta.

Covid-19 ha spinto gente curiosa, di ogni ambito scientifico e non solo degli addetti ai lavori, a fare previsioni sulla sua durata per varie aggregazioni territoriali, per età... nel perdurare dell'epidemia senza aspettarne la fine anche per motivi di organizzazione e adeguamento del sistema e delle strutture sanitarie, nonchè per gli effetti squilibranti il sistema economico sociale.

Noi abbiamo lavorato con i dati giornalieri dei decessi in Italia. Pur consapevoli che erano il risultato di esiti relativi a più focolai con tempi di insorgenza, dimensioni potenziali, terapie in atto differenti e che sarebbe stato meglio usare somme di più logistiche, abbiamo invece ripiegato sulla logistica inseguendo con questo semplice strumento non ex post ma giorno giorno le caratteristiche della pandemia Covid-19 in Italia. Abbiamo quindi ristimato ogni giorno a partire dal 21/03/2020 la miglior logistica con il dato che si aggiungeva via via: l'algoritmo che abbiamo scritto trova rapidamente il minimo della funzione obiettivo che è la somma dei quadrati degli scarti tra i dati empirici e i valori letti sulla funzione interpolante.

Queste, in TAB. 1, sono le risultanze al 15/04/2020.

Tab. 1 Stima della dinamica dei parametri di logistiche per la previsione della data di fine Covid-19 in Italia

t	a	b	c	data stima	fine Covid-19
25	0,230204	26,2450	10981,63	21-mar	17-apr
26	0,224400	27,0940	12379,45	22-mar	19-apr
27	0,226120	26,8630	11998,60	23-mar	18-apr
28	0,224520	27,0549	12300,92	24-mar	19-apr
29	0,223700	27,1433	12436,21	25-mar	19-apr
30	0,221970	27,3161	12693,78	26-mar	19-apr
31	0,215320	27,9647	13664,43	27-mar	20-apr
32	0,208220	28,6831	14775,64	28-mar	22-apr
33	0,203970	29,1265	15477,96	29-mar	23-apr
34	0,199945	29,5416	16138,03	30-mar	24-apr
35	0,196030	29,9544	16799,64	31-mar	14-apr
36	0,193067	30,2607	17288,67	01-apr	25-apr
37	0,189906	30,5752	17785,18	02-apr	26-apr
38	0,186626	30,9012	18297,13	03-apr	26-apr
39	0,183917	31,1728	18721,47	04-apr	27-apr
40	0,182225	31,3366	18973,47	05-apr	27-apr
41	0,180238	31,5260	19260,74	06-apr	28-apr

42	0,178036	31,7306	19566,34	07-apr	28-apr
43	0,175899	31,9258	19854,11	08-apr	28-apr
44	0,173380	32,1542	20186,89	09-apr	29-apr
45	0,160719	32,3950	20532,72	10-apr	29-apr
46	0,167712	32,6685	20921,60	11-apr	30-apr
47	0,165142	32,9041	21253,89	12-apr	30-apr
48	0,162357	33,1615	21612,20	13-apr	01-mag
49	0,159296	33,4489	22008,42	14-apr	01-mag
50	0,156110	33,7544	22425,48	15-apr	02-mag

I giorni di durata totale dell'epidemia sono $2b$ e le date di fine epidemia si stimano così: con i dati del 15 aprile $33.75438 \cdot 2 - 50 \simeq 17$ e si va al 2 maggio (15 aprile più 17 giorni); con i dati del 21 marzo $26.24501 \cdot 2 - 25 \simeq 27$ si sarebbe andati solo al 17 aprile ... (vedi ultima colonna).

Come si vede le stime giornaliere della logistica indicano la emidurata, b , e il numero totale dei morti, c , in crescita, anche se con qualche rallentamento: ogni $1.5 - 2.5$ giorni la stima del fine Covid-19 si sposta di 1. Si intuisce che Achille dovrebbe raggiungere la Tartaruga che ora, il 15 aprile, ha un vantaggio di 17 giorni ma Achille corre più veloce!

3. Quando Achille raggiunge la tartaruga? (Questo paragrafo si può saltare visto il modello introdotto successivamente al paragrafo 4.)

Per dare un metodo per calcolare il vero fine emergenza Covid-19, ovvero quando Achille raggiungerà la tartaruga si è proceduto basandosi sulle stime di b sopra riportate in terza colonna Tab. 1. Mediante interpolazione lineare, sempre basata sulla minimizzazione di somme di scarti quadratici. Ogni giorno, indicato con $t_* = 27, 28, \dots, 50$ (quindi con almeno tre dati di b già calcolati, poichè il primo dato è quello relativo al giorno 25) si sono stimate la pendenza $m(t_*)$ e l'intercetta $q(t_*)$ delle $24 = 50 - 27 + 1$ rette interpolanti. Questo consente di estrapolare in ogni t_* la stima di $b(t_*, t)$ (con ovvia interpretazione) per un qualsiasi t con la funzione lineare

$$b(t_*, t) \equiv q(t_*) + m(t_*) \cdot t$$

Quindi, ricordando il significato del parametro b (b è la emidurata dell'epidemia e il giorno in cui termina se l'origine dell'epidemia è posta in 0), il giorno di fine emergenza Covid-19 stimata con i dati giornalieri di decessi fino a t_* deve verificare l'equazione

$$2b(t_*, t) \equiv t$$

ovvero

$$2 \cdot (q(t_*) + m(t_*) \cdot t) = t$$

con soluzione

$$t_{sol} = \frac{2 \cdot q(t_*)}{1 - 2 \cdot m(t_*)}$$

Si riportano nella Tab.2 i valori della pendenza, $m(t_*)$, e del termine noto, $q(t_*)$, delle rette interpolanti i valori di b stimate dal giorno 27 (23-mar) al giorno 50 (15-apr), nonchè il giorno finale dell'emergenza e la data corrispondente.

Tab. 2 Interpolazione lineare per la successione delle stime del parametro b

t_*	$m(t_*)$	$q(t_*)$	t_{sol}	data corrispondente
27	0,308995	18,700133	97,9039	01-giu
28	0,219864	20,987829	74,9201	09-mag
29	0,175747	22,134871	68,2642	03-mag
30	0,162721	22,477885	66,6448	01-mag
31	0,210127	21,213715	73,1829	08-mag
32	0,272236	19,536781	85,7765	20-mag
33	0,312631	18,432651	98,3763	02-giu
34	0,338785	17,709059	109,8475	13-giu
35	0,356415	17,215417	119,8972	23-giu
36	0,364659	16,981846	125,4744	29-giu
37	0,367987	16,886418	127,9155	01-lug
38	0,368880	16,860527	128,5890	02-lug
39	0,367028	16,91484	127,2068	01-lug
40	0,36130	17,084596	123,1827	27-giu
41	0,354008	17,303542	118,5244	22-giu
42	0,346266	17,538388	114,0830	18-giu
43	0,338340	17,781460	109,993	13-giu
44	0,330993	18,009200	106,5594	10-giu
45	0,324383	18,216313	103,7280	07-giu
46	0,318816	18,392597	101,5139	05-giu
47	0,313676	18,557098	99,5960	03-giu
48	0,309155	18,703264	98,0027	02-giu
49	0,305435	18,824787	96,7535	31-mag
50	0,302510	18,921301	95,8093	30-mag

Numerosi potrebbero essere i commenti e ci limitiamo ad osservare che le stime della data di fine emergenza Covid-19, prima piuttosto erratiche, tendano a dare finalmente un'indicazione convergente guardando le previsioni per $t_* \geq 46$.

4. Modello con due focolai (ideato il 17/04/2020 e con la sperimentazione fino al 27/04/2020)

Visto che l'estrapolazione del parametro b con il modello a unico focolaio conduce molto in là la previsione della fine Covid-19, abbiamo considerato se indicazioni concordanti si avrebbero con il modello con $m = 2$ focolai, che consentirebbe di ottenere anche altre utili indicazioni. Il modello considerato è ora

$$g(t) = \frac{c_1}{1 + \exp(-a \cdot (t - b_1))} + \frac{c_2}{1 + \exp(-a \cdot (t - b_2))}$$

caratterizzato dai cinque parametri a, b_1, b_2, c_1, c_2 : trattandosi della stessa epidemia sui due focolai si è lasciato un unico parametro a e si sono distinti per i due focolai i giorni di picco, b_1 e b_2 , e i decessi totali nei due focolai, rispettivamente c_1 e c_2 . La funzione obiettivo da minimizzare diventa ora

$$\varphi(a, b_1, b_2, c_1, c_2) = \sum_{h=0}^{t_*} \left(y_h - \frac{c_1}{1 + \exp(-a \cdot (h - b_1))} + \frac{c_2}{1 + \exp(-a \cdot (h - b_2))} \right)^2$$

ovvero la somma dei quadrati degli scarti tra i dati empirici y_h (il cumulo dei decessi registrati fino al giorno h) e i valori ricavati per interpolazione. L'algoritmo di minimizzazione è un po' più complicato ma cattura la (unica) soluzione minimizzante con qualunque precisione desiderata controllando sia il valore della funzione obiettivo che quello delle 5 derivate parziali di φ : la procedura è snellita per il fatto che i valori di c_1 e c_2 minimizzanti sono determinati con formula chiusa, condizionatamente ai valori degli altri tre parametri. Si limita di seguito la presentazione dei risultati con il modello a due focolai ai giorni $t_* = 50, 51, 52, 53, 54, 55, \dots, 62$ (da 54 al 62 aggiunti nella odierna versione della nota, il 28/04/2020)

t_*	a	b_1	b_2	c_1	c_2	$c_1 + c_2$	h_1	data h_1
50	0,194181	29,7881	46,6128	15773	9210	24983	93	28-mag
51	0,192665	29,9631	47,5681	16107	9489	25596	95	30-mag
52	0,191374	30,1091	48,4217	16383	9796	26179	96	31-mag
53	0,190589	30,1937	48,8995	16538	9971	26510	97	01-giu
54	0,190249	30,2291	49,0910	16602	10038	26640	97	01-giu
55	0,189827	30,2716	49,3065	16677	10104	26781	97	01-giu
56	0,188854	30,3669	49,7678	16842	10234	27076	98	02-giu
57	0,187964	30,4516	50,1644	16987	10340	27326	99	03-giu
58	0,186909	30,5496	50,6079	17151	10450	27600	100	04-giu
59	0,185861	30,6447	51,0250	17308	10546	27853	100	04-giu
60	0,184772	30,7418	51,4394	17466	10634	28100	101	05-giu
61	0,184175	30,7941	51,6559	17550	10676	28226	102	06-giu
62	0,183502	30,8519	51,8851	17641	10713	28354	102	06-giu

Le previsioni della data di fine Covid-19, ora rappresentate nell'ultima colonna della tabella con "data h_1 ", sono state determinate via via al variare del giorno di stima t_* con la minima soluzione intera della disequazione in h

$$c_2 - \frac{c_2}{1 + \exp(-a \cdot (h - b_2))} \leq 1$$

Tale soluzione, indicata con h_1 , indica ovviamente il giorno in cui c'è ancora un decesso da registrare sul secondo focolaio. Con la stessa logica si possono determinare h_{10} e h_{100} , con ovviamente $h_{100} < h_{10} < h_1$, rispettivamente soluzioni delle disequazioni

$$c_2 - \frac{c_2}{1 + \exp(-a \cdot (h - b_2))} \leq 10$$

$$c_2 - \frac{c_2}{1 + \exp(-a \cdot (h - b_2))} \leq 100$$

Queste disequazioni risolvibili in modo immediato sono operativamente equivalenti a quelle più complicate

$$c_1 + c_2 - g(h) \leq 1 \text{ o } \leq 10 \text{ o } \leq 100$$

dato che i morti ancora da aspettarsi dopo i giorni h_1 , h_{10} e h_{100} dovuti al primo focolaio risulterebbero trascurabili rispetto a quelli del secondo focolaio: ciò lo si è verificato (dettagli omissis) calcolando

$$c_1 - \frac{c_1}{1 + \exp(-a \cdot (h_j - b_1))} \text{ con } j = 1, 10, 100$$

Da notare che nella versione del nostro studio del 19/04/2020 si preferì indicare come $t_{fine} = b_1 + b_2$, che, riletto con la simbologia appena introdotta, si colloca tra h_{100} e h_{10} .

Per completezza delle indicazioni sui plausibili giorni di fine Covid-19 si riportano i risultati nel prospetto seguente

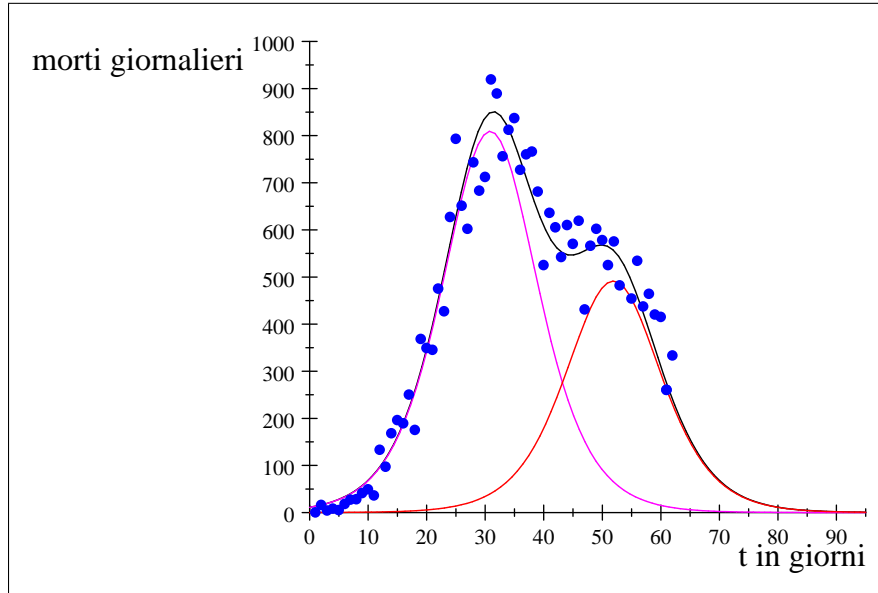
t_*	$t_{fine} = b_1 + b_2$	h_1	h_{10}	h_{100}	data h_1	data h_{10}	data h_{100}
50	76	94	82	70	29-mag	17-mag	05-mag
51	78	96	84	72	31-mag	19-mag	07-mag
52	79	97	85	73	01-giu	20-mag	08-mag
53	79	98	86	73	02-giu	21-mag	08-mag
54	79	98	86	74	02-giu	21-mag	09-mag
55	80	98	86	74	02-giu	21-mag	09-mag
56	80	99	87	75	03-giu	22-mag	10-mag
57	81	100	88	75	04-giu	23-mag	10-mag
58	81	101	88	76	05-giu	23-mag	11-mag
59	82	101	89	77	05-giu	24-mag	12-mag
60	82	102	90	77	06-giu	25-mag	12-mag
61	82	103	90	77	07-giu	25-mag	12-mag
62	83	103	90	78	07-giu	25-mag	13-mag

Consideriamo ora la funzione $g(t)$ aggiornata a $t_* = 62$: si ricorda che la derivata di $g(t)$ può in questo modello interpretarsi per $t > t_* = 62$ come previsione del numero di decessi giornalieri futuri e fino alla fine dell'emergenza, nonchè consentire di confrontare la bontà dell'approssimazione con i dati empirici fino a $t_* = 62$. Si riporta il plot in nero di $g'(t)$ con i valori dei parametri determinati in $t_* = 62$

$$g'(t) = \frac{17641 \cdot 0.183502 \cdot \exp(-0.183502 \cdot (t - 30.8519))}{(1 + \exp(-0.183502 \cdot (t - 30.8519)))^2} + \frac{10713 \cdot 0.183502 \cdot \exp(-0.183502 \cdot (t - 51.8851))}{(1 + \exp(-0.183502 \cdot (t - 51.8851)))^2}$$

quello dei suoi addendi in magenta e in rosso e i dati giornalieri dei decessi quotidiani in blu

$$\frac{17641 \cdot 0.183502 \cdot \exp(-0.183502 \cdot (t - 30.8519))}{(1 + \exp(-0.183502 \cdot (t - 30.8519)))^2}, \frac{10713 \cdot 0.183502 \cdot \exp(-0.183502 \cdot (t - 51.8851))}{(1 + \exp(-0.183502 \cdot (t - 51.8851)))^2}$$



Non credo ci siano da dare chiarimenti su come il grafico illustri in modo pressoché completo la logica del ragionamento fin qui condotto per inseguire la previsione di fine Covid-19 nel perdurare dell'emergenza. Da notare soltanto che il cresciuto valore stimato in $t_* = 62$ di b_2 , ora 51.8851, rispetto a quello, 48.8995, stimato in $t_* = 53$ ha condotto ad una $g(t)$ con due punti di massimo, assoluto il primo e relativo il secondo.

Commento finale

I due modelli ideati hanno utilizzato e continuano a utilizzare come unici dati empirici quelli della serie storica del numero dei decessi quotidiani in Italia nel perdurare dell'emergenza Covid-19. Si sono prima adattati i valori dei tre parametri, a, b, c del modello con un focolaio e dal 17 aprile i cinque parametri a, b_1, b_2, c_1, c_2 del secondo modello con due focolai. Questo ormai si impone poichè il crescere anomalo del minimo della funzione obiettivo utilizzando il modello con unico focolaio faceva intravedere che non ce l'avrebbe fatta a convergere significativamente su quello che si stava osservando. Si segnala solo che con il secondo modello i minimi della funzione obiettivo, 476 508.47 il giorno $t_* = 53$ e 698 287.28 il giorno $t_* = 62$ rappresentano quote esigue di quelli ottenibili con il primo modello: per esempio con un unico focolaio già per $t_* = 50$ il minimo, 4 559 933.48, che si poteva ottenere era circa 9.57 volte il minimo del giorno $t_* = 53$ e 6.53 volte il minimo del giorno $t_* = 62$.

Non abbiamo controllato se in letteratura esista analoga esperienza di analisi dinamica nel perdurare di epidemie (dell'attuale o di quelle passate). Ringraziamo anticipatamente se qualcuno ci indica colleghi che hanno adottato e resi noti protocolli di indagine analoghi al nostro.

Una curiosità che ci siamo tolti è stata quella di impiegare il modello a due focolai con i primi dati della serie storica per vedere da quale giorno si poteva già

prevedere l'accendersi di un nuovo significativo focolaio. Citiamo per esempio che già in $t_* = 38$, il 03-apr, le stime avrebbero dato per il modello a 2 focolai

$$a = 0,237647, b_1 = 24,4762, b_2 = 34,7062, c_1 = 11634, c_2 = 6920, c_1 + c_2 = 18554$$

con minimo funzione obiettivo paria a 116 963.42, mentre il modello a unico focolaio dava

$$a = 0.186626, b = 30.9012, c = 18297$$

con un valore già circa 5 volte più grande per il minimo della funzione obiettivo.

Guardando il grafico in corrispondenza del giorno 62, ovvero il 27-apr, il primo focolaio pare in estinzione: questo suggerisce che si potrebbero togliere i decessi giornalieri relativi al primo focolaio dai decessi quotidiani e proseguire l'indagine con le differenze riutilizzando il modello a unico focolaio e, incrociando le dita, sperando che non ci sia bisogno di reimpiego del modello a 2 focolai.

Il 04-mag è previsto In italia un primo allentamento delle rigidità imposte alla nostra vita dall'emergenza Covid-19!

Si tornerà anche fuori casa cercando di riprendere vecchie abitudini modi di vivere; diminuirà così la disponibilità di tempo da mettere a disposizione per indagini come quella da noi qui condotta con pazienza e scrupolo e sperando che la situazione evolva al meglio e non ci dia più la curiosità di aggiornarla.