
Matematica, Cultura e Società

RIVISTA DELL'UNIONE MATEMATICA ITALIANA

ANDREA PUGLIESE

Cenni su teoria ed utilizzo di modelli matematici per le epidemie

Matematica, Cultura e Società. Rivista dell'Unione Matematica Italiana, Serie 1, Vol. 5
(2020), n.1, p. 5–15.

Unione Matematica Italiana

[<http://www.bdim.eu/item?id=RUMI_2020_1_5_1_5_0>](http://www.bdim.eu/item?id=RUMI_2020_1_5_1_5_0)

L'utilizzo e la stampa di questo documento digitale è consentito liberamente per motivi di ricerca e studio. Non è consentito l'utilizzo dello stesso per motivi commerciali. Tutte le copie di questo documento devono riportare questo avvertimento.

*Articolo digitalizzato nel quadro del programma
bdim (Biblioteca Digitale Italiana di Matematica)*

SIMAI & UMI

<http://www.bdim.eu/>

Cenni su teoria ed utilizzo di modelli matematici per le epidemie

ANDREA PUGLIESE

Università di Trento

E-mail: andrea.pugliese@unitn.it

Sommario: *In questa nota divulgativa presento alcuni aspetti della teoria matematica delle epidemie, con particolare attenzione al caso di una nuova infezione emergente, come nel caso del COVID-19. Parto dalla definizione di numero di riproduzione di base R_0 , e presento la sua relazione col tasso di crescita esponenziale all’inizio di un’epidemia e con le possibilità di controllo in semplici modelli deterministici. Discuto poi i corrispondenti modelli stocastici, mostrando come in questo caso R_0 sia legato alla probabilità che esploda un’epidemia a partire dall’introduzione di pochi infetti. Infine accenno alla struttura dei modelli complessi che vengono attualmente usati come supporto previsionale alle azioni delle autorità sanitarie.*

Abstract: *In this note I present some aspects of the mathematical theory of epidemics, with particular attention to the case of a new emerging infection, as in the case of COVID-19. I start with the definition of basic reproduction number R_0 , and present its relationship with the exponential growth rate at the beginning of an epidemic and with potential for control in simple deterministic models. I then discuss the corresponding stochastic models, showing that in this case the probability that the introduction of a few infected individuals results into an epidemic outbreak depends on the value of R_0 . Finally, I will mention the structure of complex models that are currently used as support for the actions of the health authorities.*

1. – Il modello SIR di Kermack e McKendrick

I modelli attualmente usati per le epidemie in gran parte originano dal modello proposto da Kermack e McKendrick nel 1927 [12]. Anche se il modello contiene ipotesi chiaramente irrealistiche, i concetti introdotti tramite tale modello risultano essenziali per fornire una prima intuizione sulla dinamica delle epidemie, intuizione che rimane confermata in modelli più complessi, sia pure con numerose modifiche.

Tale modello prevede la divisione della popolazione in compartimenti, quello dei suscettibili, S , gli individui sani che possono essere infettati, gli infetti I ed i rimossi R , compartimento che comprende coloro che sono guariti e immuni e (nella trattazione originale) anche coloro che sono morti per la malattia.

Nella trattazione originale gli infetti vengono distinti a seconda del tempo τ trascorso dal momento dell’infezione (li chiameremo infetti con età di infezione τ); poniamo $\pi(\tau)$ la probabilità che un individuo sia ancora infetto τ istanti di tempo dopo l’infezione (in caso contrario sarà diventato un rimosso), e $\phi(\tau)$ la sua infettività (possiamo definirla come la probabilità che un suscettibile entrato in contatto con un infetto a quello stadio diventi infetto).

In teoria le funzioni $\pi(\tau)$ e $\phi(\tau)$ potrebbero essere stimate a partire da dati sperimentali per infezioni animali in laboratorio. In pratica, per infezioni umane si stimano tramite metodi statistici, a partire dall’analisi di casi in cui si riesca ad accertare il momento del contagio, due quantità collegate al periodo di infettività: il *periodo di incubazione* (ossia l’intervallo di tempo fra il momento di infezione e la comparsa di sintomi) e l’*intervallo seriale*, ossia il tempo fra la comparsa dei sintomi in un infetto e la comparsa dei sintomi in un individuo infettato dal

Accettato: il 17 marzo 2020.

primo. Per esempio per il coronavirus⁽¹⁾, Li *et al.* [13] hanno stimato, sulla base di pochissimi casi, un tempo medio di incubazione pari a 5, 2 giorni e un intervallo seriale medio di 7, 5 giorni, in entrambi i casi con intervalli di confidenza molto ampi. Nell'ultimo mese sono state ottenute numerose altre stime di questi parametri sulla base di nuovi dati disponibili [20, 21, 22].

Per completare il modello è necessario stabilire la legge secondo la quale si infettano i suscettibili. L'ipotesi fatta da Kermack e McKendrick è quella dell'*azione di massa* ovvero del "*mixing*" omogeneo; si assume cioè che ogni individuo abbia la stessa probabilità di contattare qualunque altro individuo nella popolazione, indipendentemente dai contatti passati. Si capisce come tale ipotesi sia irrealistica: siamo spesso in contatto con le persone della nostra famiglia, del nostro luogo di lavoro, classe scolastica, gruppo di amici e molto di rado con chi abita in un luogo diverso, ha età e professioni diverse. Negli ultimi anni si sono quindi sviluppati modelli computazionali (descritti in seguito) che cercano di tenere conto di molte caratteristiche degli individui per arrivare a previsioni più accurate; però la semplicità della legge di azione di massa permette di effettuare un'analisi teorica precisa e quindi di raggiungere conclusioni potenti.

Supponiamo allora che ogni individuo, e in particolare ogni suscettibile, contatti in media β individui nell'unità di tempo e si vuole modellizzare l'*incidenza* $i(t)$, ossia il tasso di nuove infezioni; in altre parole il numero di nuove infezioni nell'intervallo infinitesimo $(t, t + dt)$ è dato da $i(t) dt$. Al tempo t gli individui con età di infezione τ sono $i(t - \tau)\pi(\tau)$: coloro che si sono infettati al tempo $t - \tau$ e sono ancora infetti. La probabilità che un individuo contattato sia un infetto con età di infezione τ è quindi pari a $\frac{i(t - \tau)\pi(\tau)}{N(t)}$, dove $N(t)$ è la popolazione totale; infine $\phi(\tau)$ è la probabilità di infettarsi a seguito di contatto con un infetto con età di infezione τ .

Integrando su tutte le possibili età di infezione (che per comodità matematica supponiamo possa-

no essere qualunque numero positivo) otteniamo l'equazione

$$(1) \quad i(t) = S(t)\beta \int_0^\infty \frac{i(t - \tau)\pi(\tau)\phi(\tau)}{N(t)} d\tau.$$

Per completare il sistema, supponiamo che l'epidemia si svolga su un tempo sufficientemente breve da poter ignorare nascite e morti per altre cause, e che anche le morti causate dalla malattia siano sufficientemente poche rispetto alla popolazione totale, da poter supporre che $N(t)$ sia una costante N . Quindi $S(t)$ cambia solamente a causa di nuove infezioni e otteniamo il sistema

$$(2) \quad \begin{aligned} S'(t) &= -i(t) \\ i(t) &= \beta \frac{S(t)}{N} \int_0^\infty i(t - \tau)\pi(\tau)\phi(\tau) d\tau. \end{aligned}$$

La seconda equazione presuppone che l'infezione proceda da un tempo infinito, e ciò porrebbe problemi sia biologici sia matematici. Si può supporre invece che l'infezione inizi al tempo 0 con un certo numero di infetti di diverse età di infezione $i_0(\tau)$; si può allora trasformare (2) in

$$(3) \quad \begin{aligned} S'(t) &= -i(t) \\ i(t) &= \beta \frac{S(t)}{N} \left(\int_0^t i(t - \tau)\pi(\tau)\phi(\tau) d\tau + \right. \\ &\quad \left. + \int_t^\infty \frac{\pi(\tau)}{\pi(\tau - t)} \phi(\tau) i_0(\tau - t) d\tau \right) \end{aligned}$$

dove il secondo integrale tiene conto delle infezioni causate al tempo t da coloro che erano già infetti al tempo 0. Supponendo che sia nota la funzione $i_0(\tau)$ e il valore di $S(0) = S_0$, (3) è un problema ben posto sotto ipotesi minime (e biologicamente ragionevolissime) sulle funzioni i_0 , ϕ e π ; i lettori interessati possono trovare le dimostrazioni di questo e dei punti successivi in [11].

2. – Crescita esponenziale e numero di riproduzione di base R_0

L'analisi dei sistemi (2) o (3) necessita di strumenti avanzati, ma la conclusione sull'andamento iniziale dell'epidemia è relativamente semplice e conviene usare (2). All'inizio dell'epidemia gli infetti e le nuove

⁽¹⁾ il nome ufficiale è SARS-CoV2 per il virus e COVID-19 per la malattia causata, ma nell'articolo verrà chiamato semplicemente coronavirus

infezioni sono poche e si può quindi approssimare $S(t) \approx S_0$, ottenendo un'equazione lineare per $i(t)$:

$$(4) \quad i(t) = \beta u_0 \int_0^\infty i(t-\tau) \pi(\tau) \phi(\tau) d\tau.$$

dove $u_0 = S_0/N$. Nel caso di una nuova infezione, non vi è immunità pregressa e quindi $S_0 \approx N$, ossia $u_0 \approx 1$. Si dimostra che le soluzioni di (4) crescono (o decrescono) esponenzialmente, $i(t) \approx Ce^{rt}$. Sostituendo questa espressione in (4), si nota subito che r deve risolvere l'equazione trascendente

$$(5) \quad \beta u_0 \int_0^\infty \pi(\tau) \phi(\tau) e^{-r\tau} d\tau = 1.$$

Si dimostra facilmente che (5) ha un'unica soluzione reale; inoltre $r > 0$ [ovvero $r < 0$], ossia l'epidemia cresce [ovvero decresce] se e solo se $R_0 u_0 > 1$ [ovvero $R_0 u_0 < 1$] dove

$$(6) \quad R_0 = \beta \int_0^\infty \pi(\tau) \phi(\tau) d\tau.$$

R_0 è detto il *numero di riproduzione di base dell'epidemia* e rappresenta il numero medio di individui infettati da un infetto nel corso del suo periodo infettivo, supponendo che ogni altro individuo nella popolazione sia suscettibile. Infatti, se pensiamo a un individuo infettato al tempo 0, esso sarà ancora infetto al tempo τ con probabilità $\pi(\tau)$, a quel tempo contatterà altri individui (necessariamente suscettibili) al tasso β e li infetterà con probabilità $\phi(\tau)$. Integrando su tutti i tempi τ da 0 a ∞ , si trova il numero totale di individui infettati lungo tutto il periodo infettivo.

Se invece u_0 è la proporzione di popolazione suscettibile, il numero totale infettato sarà $R_0 u_0$.

La condizione $R_0 u_0 > 1$ perché un'epidemia si possa diffondere (il cosiddetto *teorema di soglia* di Kermack e McKendrick) diventa allora molto intuitiva. Se $R_0 u_0 > 1$, ogni individuo infetto ne infetta in media più di 1, che a sua volta ne infetterà più di 1, e così via, con una reazione a catena esplosiva. Se invece $R_0 u_0 < 1$, ogni individuo infetto ne infetta in media meno di 1, e quindi l'epidemia si estinguerà velocemente da sola.

L'importanza della quantità R_0 è chiara anche ai fini del controllo. Se avessimo a disposizione un vaccino (perfetto, ossia che fornisce immunità totale a tutti i vaccinati) dovremmo vaccinare abbastanza

persone da far sì che la quantità $R_0 u_0$ diventi minore di 1. Se p fosse la percentuale della popolazione vaccinata, u_0 sarebbe uguale a $1 - p$ (la percentuale non vaccinata). Per eliminare un'infezione dovremmo quindi avere $R_0(1 - p) < 1$, ossia la percentuale minima di individui da vaccinare per eliminare un'infezione è pari a $1 - \frac{1}{R_0}$. Sono state ottenute stime di R_0 per varie infezioni; per esempio le stime di R_0 per il vaiolo erano fra 5 e 7, così come quelle per la polio; invece le stime di R_0 per il morbillo variano fra 12 e 18. Questo spiega il target del 95% che ci si pone per la vaccinazione del morbillo, e spiega anche perché sia più facile eradicare il vaiolo (il grande successo delle campagne di vaccinazione) e la polio (si è arrivati molto vicini al successo) che il morbillo.

Conoscere il valore di R_0 è utile per impostare strategie di controllo, anche se non si ha a disposizione un vaccino e si vuole intervenire tramite la riduzione del tasso di contatto. Se per esempio R_0 fosse uguale a 2, dovremmo più che dimezzare i tassi di contatto per controllare l'epidemia; se fosse uguale a 3, la riduzione dovrebbe essere proporzionalmente maggiore. Per l'infezione da coronavirus, le stime di R_0 fatte sulla base dei dati osservati in Cina [18] variano in genere fra 2 e 3, e quelle fatte sulla base delle prime 2-3 settimane di dati italiani sono anche più alte. Questo può dare un'idea dello sforzo necessario per il suo contenimento.

Ma come si sono ottenute queste stime di R_0 ? In sostanza si sono trovate inserendo (6) all'interno di (5) e ottenendo (supponendo inoltre $u_0 = 1$)

$$(7) \quad R_0 = \frac{\int_0^\infty \pi(\tau) \phi(\tau) d\tau}{\int_0^\infty \pi(\tau) \phi(\tau) e^{-r\tau} d\tau}.$$

Se conosciamo le funzioni $\pi(\tau)$ e $\phi(\tau)$ (o qualche loro approssimazione) e abbiamo stimato r , ne otteniamo una stima di R_0 . Stimare r è in realtà abbastanza semplice se possiamo osservare la curva dei nuovi casi $i(t)$; supponendo, come detto sopra, che all'inizio dell'epidemia si abbia $i(t) \approx Ce^{rt}$, possiamo stimare r con una regressione lineare di $\log(i(t))$ su t . Un vantaggio di questo metodo è che funziona anche se non vengono riportati tutti i casi ma solo una frazione p (per esempio quelli ospedalizzati perché più gravi); purché il fattore di sottotifica p rimanga costante nel tempo, la stima di r rimane non distorta.

Metodi più elaborati per stimare i parametri epidemiologici si basano sull'analisi delle sequenze geniche dei virus e la ricostruzione degli alberi di trasmissione; si è sviluppata una branca statistica, nota come filodinamica [16], dedicata a questi problemi, ma ciò esula dal livello di questo articolo.

Parlando di stime dei parametri epidemiologici all'inizio di un'epidemia, vale la pena chiarire perché è difficile ottenere una buona stima di una quantità che è certamente molto importante, come la letalità di un'infezione. Si potrebbe pensare che basti calcolare il rapporto fra il numero dei morti e quello dei casi registrati a un certo tempo t per avere una stima della probabilità di morire a causa dell'infezione. Al di là del fatto che la mortalità dipenderà dalle condizioni dell'infettato, e in particolare dalla sua età (per il coronavirus ciò è molto evidente sia in Cina [19] sia in Italia, secondo analisi preliminari dell'Istituto di Sanità), la stima della mortalità tramite il rapporto fra il numero dei morti e quello dei casi registrati soffre di due problemi. Prima di tutto, mentre il numero dei morti è in genere abbastanza certo, i casi registrati sono in genere solo una parte delle infezioni effettivamente avvenute. Il secondo problema è che in genere il tempo di degenza ospedaliera fino alla morte può essere più o meno lungo; quindi al tempo t possiamo conoscere chi è morto al tempo t , ma non sappiamo quanti fra i casi già registrati al tempo t moriranno per la malattia nei prossimi giorni o settimane. Il primo fattore ci fa sovrastimare la mortalità, il secondo, specialmente in un periodo di crescita esponenziale dei casi, ce la fa sottostimare. Quando ci sono tanti dati disponibili, è possibile effettuare correzioni adeguate; in una prima fase, invece, le stime della mortalità non sono molto attendibili.

3. – Il modello SIR con equazioni differenziali ordinarie

La formula che collega R_0 e r diventa molto più semplice, se supponiamo che l'infettività ϕ sia costante e che il periodo di infezione segua una distribuzione esponenziale $\pi(\tau) = e^{-\gamma\tau}$. Allora (7) diventa

$$(8) \quad R_0 = 1 + \frac{r}{\gamma} = 1 + rT_I$$

dove T_I rappresenta la lunghezza media del periodo di infettività. La formula (8) si usa spesso come approssimazione anche nel caso in cui la distribuzione del periodo di infettività non sia esponenziale, anche, se come visto sopra, ciò non è corretto. Relazioni esplicite si possono ottenere anche nel caso si scelgano distribuzioni diverse dall'esponenziale.

L'ipotesi che il periodo di infettività segua la distribuzione esponenziale, equivale a supporre che la probabilità di guarire non dipenda da quanto a lungo un individuo è stato infetto; come è intuitivo, in genere le distribuzioni effettive sono molto lontane dall'esponenziale. Però la distribuzione esponenziale ha il grosso vantaggio di trasformare il sistema (2) consistente in un'equazione differenziale e un'equazione integrale accoppiate in un sistema di equazioni differenziali ordinarie che sono molto più semplici da analizzare e da approssimare al computer. Il sistema che si ottiene è

$$(9) \quad \begin{aligned} S'(t) &= -\beta\phi\frac{S(t)}{N}I(t) \\ I'(t) &= \beta\phi\frac{S(t)}{N}I(t) - \gamma I(t). \end{aligned}$$

dove

$$I(t) = \int_0^\infty i(t-\tau)e^{-\gamma\tau} d\tau$$

rappresenta il numero totale di infetti. Questo modello è noto come il modello SIR, perché gli individui partono tutti dallo stato S, una parte di loro entra in I e infine in R.

Volendo andare un po' più vicino alla realtà, pur rimanendo nell'ambito delle equazioni differenziali ordinarie, si può introdurre una classe intermedia, E , di individui che sono stati infettati ma non sono ancora infettivi, ottenendo il modello SEIR. Il periodo che gli individui passano nello stadio E si può considerare come il periodo di incubazione, anche se nella realtà è possibile che gli individui infetti siano in grado di trasmettere l'infezione prima di manifestare sintomi.

Come abbiamo detto, le soluzioni del sistema (9), così come quelle di (2) hanno una prima fase di crescita esponenziale a tasso r , positivo purché $R_0 u_0 > 1$, come si suppone in genere (altrimenti non si osserverebbe l'epidemia).

Anche modelli più complessi, in cui, come accennato, la popolazione viene stratificata in base a molte

altre caratteristiche, hanno questa proprietà e le relazioni fra r e R_0 sono simili a quanto visto in (7).

Nel caso di modelli in cui la popolazione è strutturata in gruppi con diverse caratteristiche, un metodo generale per la definizione ed il calcolo di R_0 è stato proposto da Diekmann *et al.* [6].

4. – Il tasso finale di attacco di un’epidemia

Chiaramente la fase di crescita esponenziale non può andare avanti per sempre, anche qualora non vengano messe in atto strategie di controllo o mitigazione. Infatti il numero di suscettibili $S(t)$ non potrà che diminuire (non sono previsti ingressi di nuovi suscettibili) e a un certo punto si arriverà a $R_0 \frac{S(t)}{N} = 1$; a quel punto, si sarà raggiunto il picco di infetti, e l’epidemia comincerà a declinare, finché (in un tempo infinito, secondo il modello matematico in cui il numero di infetti può raggiungere frazioni minime, ma non può mai diventare uguale a 0) non ci saranno più infetti, mentre sarà rimasta una certa frazione ancora suscettibile. La frazione rimasta suscettibile alla fine dell’epidemia secondo i modelli (9) o (2) si può calcolare esattamente, e dipende solo da R_0 ; nel pannello di destra della Figura 1 sotto si mostra come il tasso di attacco (ossia la frazione della popolazione che è stata infettata) dipenda da R_0 , supponendo che non vi sia immunità iniziale.

Si vede come già per $R_0 = 2$ il modello preveda che circa l’80% della popolazione venga infettata nel corso di un’epidemia.

Però la previsione ottenuta tramite questo modello è quasi certamente una sovrastima di quanto ci si possa aspettare in realtà. Prima di tutto, modelli più complessi in cui si tiene conto dell’eterogeneità della popolazione prevedono quasi sempre, a parità di R_0 o di r , un tasso di attacco più basso di quello del modello con “mixing” omogeneo, anche se questa proprietà si riesce a dimostrare solo per alcune classi di modelli. Inoltre, al di là di eventuali strategie di prevenzione imposte o suggerite dalle autorità, è probabile che, proprio a causa della paura indotta da un’infezione poco conosciuta, cambino i comportamenti individuali, riducendo i contatti. Una parte della teoria delle epidemie, la cosiddetta epidemiologia comportamentale, cerca di modellizzare questo tipo di fenomeni; il libro edito da Manfredi e D’Onofrio [14] fornisce un quadro completo delle ricerche in questo campo, mentre [8] ne mostra le applicazioni alle strategie di vaccinazione. Mi sembra difficile che questi modelli riescano ad essere predittivi, ma certamente comportamenti di auto-prevenzione tendono a ridurre la velocità di diffusione di un’epidemia e anche il suo tasso di attacco.

Il cenno ai comportamenti individuali è l’occasione per ricordare che il valore di R_0 in parte dipende dalla biologia del virus e delle sue interazioni

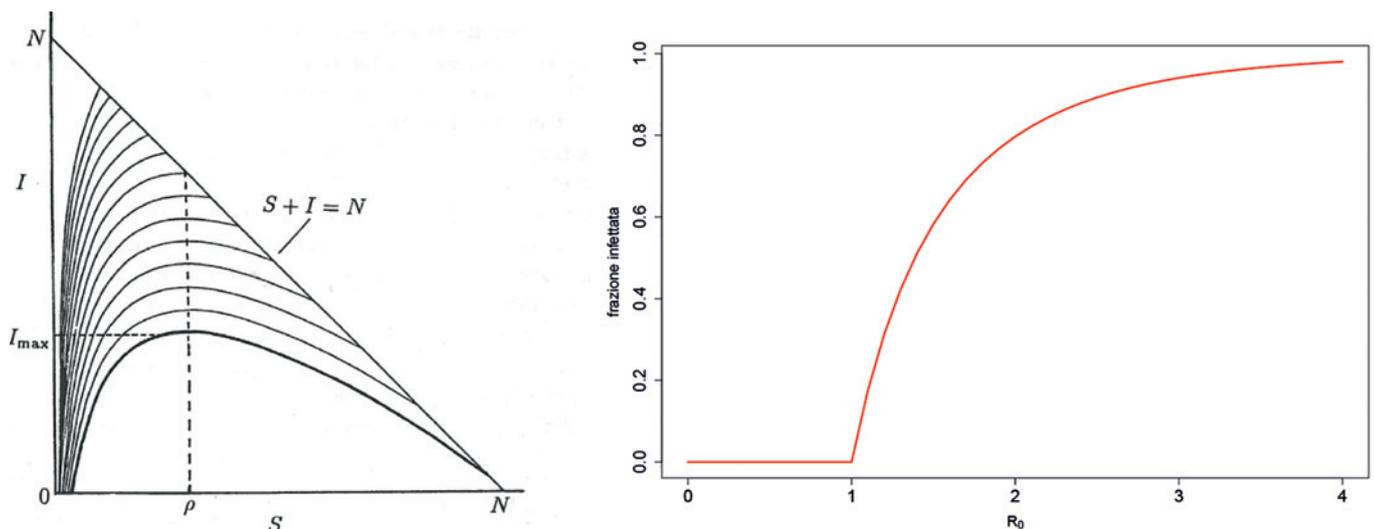


FIGURA 1 – Nel pannello di sinistra, alcune soluzioni del modello deterministico per $R_0 = 3$ con diversi valori iniziali nel piano (S, I) ; le traiettorie vanno da destra verso sinistra. Nel pannello di destra il tasso finale di attacco, ossia la frazione della popolazione infettata, in funzione di R_0 .

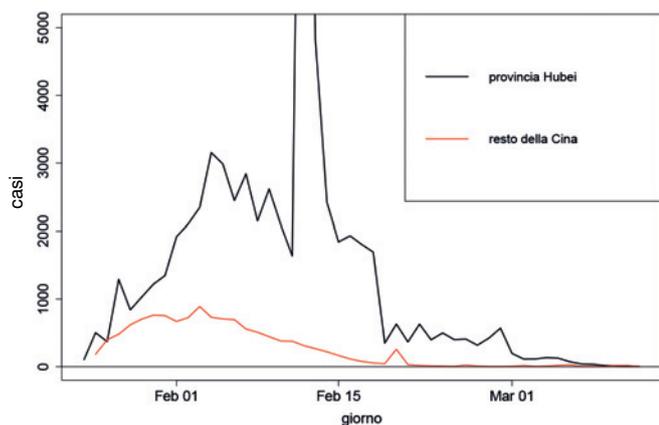


FIGURA 2 – I casi giornalieri riportati nella provincia di Hubei e nel resto della Cina fra il 24 gennaio e il 26 febbraio. Il grande picco a Hubei il 12 febbraio è dovuto ad una revisione dei criteri per la definizione di un caso.

col sistema immunitario che ne determinano l'infettività e la durata del periodo infettivo, ma in parte dipende dalla struttura dei contatti fra gli individui. Mentre le stime ottenute in Cina sul periodo di incubazione e sull'intervallo seriale dovrebbero essere approssimativamente valide dovunque, non è detto che il valore di R_0 debba essere uguale in Cina, in Italia o in Nuova Zelanda.

La dinamica dell'epidemia da coronavirus in Cina in cui da inizio febbraio vi è una discesa costante del numero giornaliero di nuovi casi accertati (vedi Figura 2) mostra che misure draconiane di isolamento sono effettivamente in grado di controllare un'epidemia. Scrivo l'ultima versione di questo articolo un paio di giorni dopo l'entrata in vigore in Italia di misure dello stesso tipo, ma senza poterne conoscere l'effetto; mi sento di essere ragionevolmente ottimista sulla loro efficacia, sperando di non essere smentito dai fatti.

La domanda naturale che sorge è quanto a lungo sia necessario mantenere in vigore tali controlli. Se usiamo il modello (9) o comunque qualsiasi modello deterministico in cui il numero di infetti sia descritto tramite una variabile continua, la risposta è "per sempre": nel modello, se sospendiamo le misure al tempo T quando il numero di infetti previsto $I(T)$ è uguale, per esempio, a 10^{-20} , l'epidemia riparte dopo il tempo T , raggiungendo un picco simile a quello che si sarebbe ottenuto inizialmente senza misure di controllo. È chiaro che in questo caso il modello sta fornendo previsioni insensate; in realtà

quando l'ultimo infetto sarà guarito, il virus sarà eradicato e l'epidemia non potrà ripartire, così come è avvenuto per la SARS nel 2003.

5. – I modelli stocastici

Per studiare dal punto di vista matematico i fenomeni di estinzione (dell'epidemia o di popolazioni in genere) è necessario passare ad un modello stocastico, in cui il numero di infetti viene descritto non tramite una variabile continua, ma da un numero intero, ed in cui le transizioni (infezioni, guarigioni, ...) fra stati avvengono in modo casuale. È possibile ottenere risultati pressoché completi anche per un modello stocastico equivalente al sistema (2) che ammette una distribuzione arbitraria del periodo di infezione ed infettività variabile nel corso di esso.

Per questo articolo preferisco limitarmi al caso più semplice di distribuzione esponenziale ed infettività costante, in cui il sistema può essere descritto come un processo di Markov nelle variabili $S(t)$ e $I(t)$ che possono assumere valori negli interi nonnegativi tali che $S(t) + I(t) \leq N$.

Le transizioni fra gli stati dipendono soltanto dallo stato attuale (questa è la proprietà di Markov) e sono analoghe a quelle previste nel sistema (9): la probabilità che ci sia una nuova infezione nell'intervallo infinitesimo $(t, t + dt)$ è pari a $\beta \frac{S(t)}{N} I(t) dt$ e la probabilità che in tale intervallo ve ne sia più di una è trascurabile; analogamente per le guarigioni.

Quando il numero di individui N della popolazione è basso, si può facilmente risolvere il sistema lineare che determina le probabilità $p_{ij}(t)$ che vi siano i suscettibili e j infetti al tempo t per tutti i valori possibili di i e j . Quando N è grande (e in molti casi basta che N sia intorno a 1.000 o addirittura 100 individui) si può ottenere una descrizione accurata della dinamica in modo semplice.

Il punto essenziale è che all'inizio dell'epidemia (quando gli infetti sono pochi) si può approssimare la dinamica con quella di un processo di ramificazione, ossia un processo in cui gli individui vivono per un tempo casuale (indipendentemente l'uno dall'altro) e durante la vita producono figli secondo una certa legge casuale. In questo caso gli individui corri-

spondono agli infetti e i figli alle nuove infezioni. Se il numero medio di figli prodotti da ogni individuo, R_0 , è minore di 1, allora il processo di ramificazione si dice *subcritico* e si ha estinzione certa della popolazione in un tempo breve, con la generazione di pochi discendenti del primo individuo. Se invece $R_0 > 1$, il processo si dice *supercritico* ed è possibile che la popolazione cresca all'infinito, ma è anche possibile che si estingua velocemente.

Nel caso di un processo di Markov (distribuzioni esponenziali), la probabilità di estinzione è $1/R_0$; per esempio, se $R_0 = 2$, c'è il 50% di probabilità che introducendo un individuo infetto in una popolazione, l'epidemia si estingua velocemente da sola, dopo che si è verificato un numero molto piccolo di infezioni.

Se invece l'epidemia non si estingue velocemente, allora il numero di infetti continuerà a crescere e l'approssimazione tramite il processo di ramificazione non sarà più valida. Quando il numero di infetti è elevato, invece, si può approssimare, grazie alla legge dei grandi numeri, il processo stocastico con le soluzioni del modello deterministico (9).

Riassumendo, si può dire che, quando $R_0 > 1$ e un individuo infetto viene introdotto in una popolazione, possono succedere due cose: o, con probabilità $1/R_0$, l'epidemia si estingue da sola con pochi casi, oppure, con probabilità $1 - \frac{1}{R_0}$, si ha una grande epidemia

nella quale il numero dei casi segue all'incirca quanto previsto dal modello deterministico. Nella Figura 3 mostro nel pannello di sinistra alcune simulazioni del modello stocastico confrontate con la soluzione del modello deterministico; nel pannello di destra la distribuzione del numero totale di casi: si vede che si tratta di una distribuzione bimodale: o ci sono pochissimi casi oppure il numero dei casi è vicino alle previsioni del modello deterministico, mentre valori intermedi sono pressoché impossibili.

I calcoli precedenti sono stati effettuati supponendo che venga introdotto un singolo infetto; chiaramente, la probabilità che si abbia una grande epidemia aumenta se si ha una ripetuta introduzione di infetti, provenienti, ad esempio, da una regione in cui l'epidemia è in corso.

Una possibilità di controllo molto praticata è il "contact tracing", consistente, una volta che sia stato identificato un individuo infetto, nel controllare tutti i suoi contatti e nell'isolarli qualora mostrino sintomi. Un recente studio di Hellewell *et al.* [9] ha calcolato la probabilità di riuscire a contenere tramite il "contact tracing" un'epidemia con caratteristiche simili a quelle del coronavirus.

Sono state effettuate analisi di scenario variando il valore di R_0 , la probabilità di trovare i contatti di un caso, l'eterogeneità fra gli individui nella trasmissione potenziale, la proporzione di infezioni che

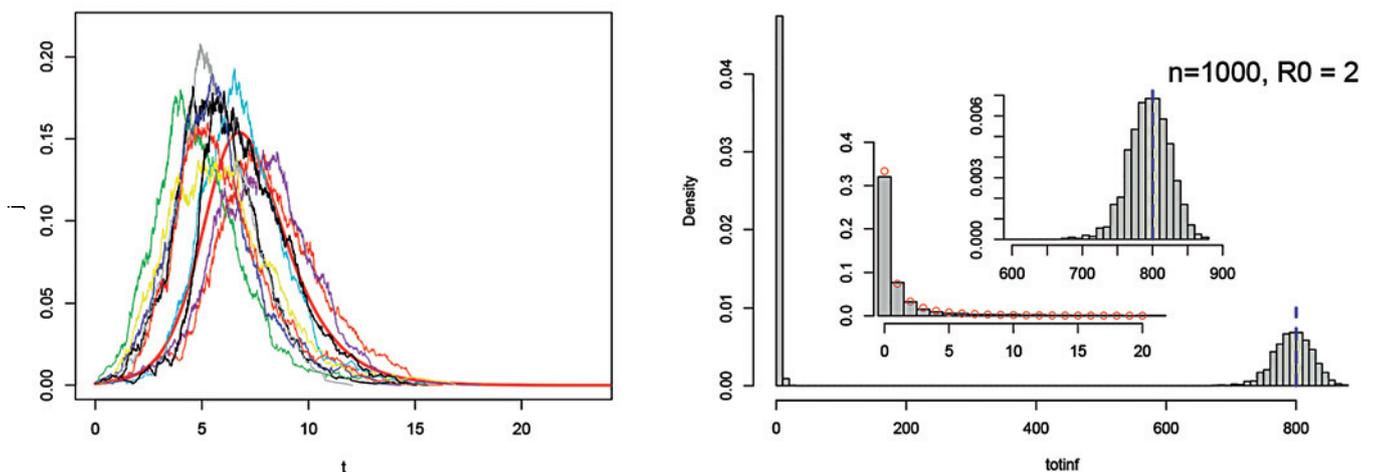


FIGURA 3 – Nel pannello di sinistra, alcune simulazioni del processo di Markov che partono da 1 infetto iniziale e 999 suscettibili; R_0 è uguale a 2; la linea rossa continua è la soluzione del modello deterministico con gli stessi parametri; non vengono mostrate simulazioni caratterizzate da estinzione rapida. Nel pannello di destra, la distribuzione finale del numero di casi osservati in 2.000 simulazioni con parametri come a sinistra; la linea tratteggiata azzurra indica il numero atteso secondo il modello deterministico. Per mostrare meglio la distribuzione, gli inserti mostrano un ingrandimento intorno allo 0 (i pallini rossi indicano un'approssimazione con la distribuzione geometrica) ed intorno al valore previsto dal modello deterministico (dove vale un'approssimazione, non mostrata, con la distribuzione normale).

avvengono quando l'individuo infetto è (ancora) asintomatico.

I risultati trovati possono essere riassunti dicendo che il successo nel contenere una tale epidemia richiede in genere di individuare e controllare almeno l'80% dei contatti, a meno che R_0 sia basso (intorno a 1,5); il contenimento diventa ancora più difficile se una parte significativa delle trasmissioni di infezione avviene nella fase asintomatica dell'infezione.

Il risultato teorico è stato purtroppo confermato dal fatto che, nonostante i controlli effettuati sui viaggiatori provenienti dalla Cina e sui loro contatti, si è sviluppata l'epidemia di coronavirus in Europa ed America. Un sistema di sorveglianza centrato sull'effettuare test solo alle persone con contatti accertati con la Cina non poteva funzionare con un virus con trasmissione anche in fase asintomatica o pre-sintomatica.

6. – I modelli ad agente individuale

Come accennato all'inizio, negli ultimi anni, grazie alla potenza di calcolo disponibile, si è diffuso l'uso di modelli ad agente individuale. Modelli di questo tipo, il cui prototipo è dovuto a Ferguson *et al.* [7] sono attualmente usati come supporto previsionale alle azioni del governo italiano nell'emergenza dell'epidemia da coronavirus (così come lo furono nella pandemia influenzale del 2009 [3]), in quanto, se idoneamente parametrizzata, possono essere più accurati di modelli basati sull'azione di massa, ma soprattutto permettono di fornire previsioni (sia pure con un'ampia banda di incertezza) sull'effetto di azioni specifiche come la chiusura di scuole o uffici in particolari contesti geografici.

In un modello ad agente individuale viene creata una popolazione artificiale che riproduca, dal punto di vista statistico, le proprietà della popolazione che si vuole studiare. Per esempio gli individui vengono distribuiti nei comuni italiani come nella popolazione reale; vengono create 'famiglie' con una distribuzione di componenti e delle loro età simile alle statistiche ufficiali; sempre seguendo le distribuzioni statistiche vengono localizzate 'scuole' e 'luoghi di lavoro' a cui gli individui vengono assegnati secondo algoritmi che tengono conto di età e distanza; a se-

conda del dettaglio, si potrebbero aggiungere mezzi di trasporto pubblico, luoghi di ritrovo, ecc. A ognuno di questi ambienti ('famiglie', 'scuole'...) viene attribuito uno specifico tasso di contatto (che può dipendere dal numero di appartenenti) fra coloro che lo frequentano, contatti che possono far trasmettere l'infezione da un suscettibile a un infetto; infine si aggiunge sempre una quota residua di contatti secondo la legge di azione di massa, che rappresenta tutte le interazioni sociali che non sono state specificamente modellizzate. È chiaro che un modello di questo tipo richiede, oltre a grosse capacità di calcolo, anche un gran numero di parametri.

Dal punto di vista matematico, il modello consiste in una catena di Markov (a tempo continuo o, più spesso, discreto) con un grandissimo numero di stati: a^N dove N è il numero di individui (circa 60 milioni per l'Italia) e a è il numero di stati epidemici. Tipicamente a sarà intorno a 10 o più, perchè ogni individuo potrà essere suscettibile, infettato ma non ancora infettivo, infettivo e sintomatico, infettivo e asintomatico, ospedalizzato, guarito; inoltre spesso ogni classe epidemiologica viene divisa in stadi in modo da approssimare meglio distribuzioni realistiche che non sono esponenziali.

Oltre ad avere un grandissimo numero di stati, si tratterà in genere di una catena senza particolari simmetrie, in quanto ogni individuo ha interazioni diverse dagli altri. Non sembra quindi possibile raggiungere risultati analitici per tali modelli e ci si affida ai risultati di un gran numero di simulazioni, sulla base delle quali stimare le probabili dinamiche epidemiche.

In modelli di questo tipo è probabile che si formino cluster di casi, per esempio quando l'infezione si diffonde in una o due 'scuole' per poi propagarsi attraverso i contatti familiari. Guardando invece le simulazioni da una scala macroscopica, tipicamente mostrano una prima fase di crescita tipo esponenziale a un tasso abbastanza simile fra una simulazione e l'altra. È quindi possibile attribuire, in modo semi-empirico, un valore di R_0 anche a un modello di questo tipo, usando la formula (7) con il valore di r stimato dalla crescita esponenziale. Inoltre nelle simulazioni si sa esattamente chi ha infettato chi; quindi si può calcolare il numero di individui infettati da ogni infetto e calcolarne la

media nel corso dell'epidemia; anche se tale numero sarà abbastanza variabile, nelle fasi iniziali di un'epidemia si avvicina a quanto si calcola tramite (6). L'eterogeneità estrema della popolazione in un modello di questo tipo fa sì che la proporzione infettata alla fine dell'epidemia sia invece in genere sensibilmente più bassa di quella di un modello basato sulla legge dell'azione di massa con lo stesso valore di R_0 [2].

Come detto sopra, non sembra possibile ottenere risultati analitici su modelli ad agente individuali. Sono stati invece sviluppati modelli che cercano di tenere conto di alcune di queste strutture sociali, ma mantenendo una struttura relativamente semplice che permette di ottenere risultati rigorosi. In [4] si suppone che la popolazione consista di N 'famiglie', ognuna delle quali con un certo numero di componenti; ogni individuo avrà contatti con ogni altro individuo della popolazione a un dato tasso, mentre con ogni individuo della stessa 'famiglia' i contatti saranno più frequenti; si può dire che ci sia una legge di azione di massa a due livelli. Similmente ai modelli con 'mixing' omogeneo, si riesce a trovare una condizione (asintotica, cioè valida per N sufficientemente grande) di soglia, sopra la quale è possibile lo sviluppo di una grande epidemia, e si sono anche definiti vari concetti corrispondenti a R_0 .

Questa idea è stata estesa a modelli più complessi in cui si suppone che ci siano 'famiglie' e 'luoghi di lavoro' e che ogni individuo appartenga a una 'famiglia' e a un 'luogo di lavoro'. Anche per questo modello si è arrivati a definire la condizione di soglia per l'insorgere di epidemie, e ad ottenere varie possibili definizioni di R_0 [15].

7. – Infezioni endemiche

Finora abbiamo considerato un'epidemia su una scala breve di tempo, ignorando nascite e morti. In ogni caso l'epidemia si estingue perché i suscettibili non sono più sufficienti a sostenerla, ossia $R_0 \frac{S(t)}{N}$ diventa minore di 1 per t grande. Cosa succede se consideriamo le nascite (oltre alle morti) che porteranno nuovi suscettibili nel sistema?

Nei modelli deterministici la risposta è abbastanza semplice: l'infezione diventerà endemica, con

le frazioni suscettibili e immuni che, nei modelli più semplici, tenderanno a un equilibrio [10]; in modelli più complessi, potrà anche succedere, a seconda delle ipotesi fatte, che le frazioni suscettibili e infette abbiano oscillazioni regolari o anche irregolari (caotiche) in alcuni casi [17, 1].

Nei modelli stocastici la risposta è meno chiara: può succedere che, dopo il picco dell'infezione e la successiva discesa, il numero di infetti arrivi a 0, prima che il numero di suscettibili sia sufficientemente cresciuto, e a quel punto l'infezione sarà estinta; la probabilità di un tale evento può essere approssimata con metodi diversi [5]. Può succedere invece che l'infezione riesca a persistere a livelli molto bassi fino a un secondo picco epidemico, che potrebbe essere seguito a sua volta dall'estinzione; oppure l'infezione può sfuggire all'estinzione e diventare endemica, oscillando per tempi lunghi fra valori vicini all'equilibrio deterministico.

Certamente la dimensione N della popolazione gioca un ruolo importante nel determinare la persistenza di un'infezione. Ciò ha riscontro anche in dati empirici: per esempio, nel periodo pre-vaccinale il morbillo non era in grado di persistere in Islanda (circa 100.000 abitanti): per molti anni non vi erano casi di morbillo sull'isola, poi succedeva che da una nave sbarcava qualche infetto, da cui scaturiva una grossa epidemia su tutta l'isola; nel giro di pochi mesi l'epidemia si esauriva e l'Islanda rimaneva libera del morbillo fino alla successiva introduzione.

La situazione diventa molto più complessa nel caso non si abbia a che fare con una singola popolazione isolata, ma l'infezione si possa diffondere, anche in modo non sincronizzato, fra varie popolazioni. Una delle domande che ci si pone adesso riguardo al coronavirus SARS-CoV2 è se, dopo questa epidemia, esso continuerà a circolare nel mondo. Nessuno ha adesso gli elementi per potere fornire una risposta, ma certamente il rischio della sua persistenza è uno scenario da considerare.

Ciò pone un problema per le strategie di isolamento sociale; la Cina è riuscita a bloccare l'epidemia e a far scendere le nuove infezioni a numeri molto bassi, ed è possibile che, se la Cina fosse isolata, si riuscirebbe a eliminarne il virus, e quindi a potere completamente revocare le misure di controllo.

Invece la Cina non è isolata e la grande maggioranza dei suoi cittadini sono ancora suscettibili all'infezione, proprio perché la strategia di controllo è stata efficace. Anche dopo l'emergenza sarà quindi necessario continuare ad effettuare test su malati con sintomi di tipo respiratorio ed essere pronti ad effettuare nuovi interventi di controllo. Proprio il successo nella strategia di controllo lascia la Cina in uno stato potenzialmente instabile, almeno fino a che il virus non venga eradicato su tutto il pianeta o non sia pronto un vaccino. Naturalmente gli stessi problemi si potranno porre in Italia nel caso favorevole in cui l'epidemia venga bloccata su numeri solo un po' più alti degli attuali.

Concludendo, ho cercato di fare un breve riassunto di alcuni concetti che emergono dalla modellizzazione matematica delle epidemie e di come possono aiutare anche nel comprendere la situazione nel caso di nuove infezioni e nel formulare strategie di controllo adeguate.

Ho anche cercato di mostrare alcuni dei problemi matematici che nascono da questi modelli, con fenomeni ancora non completamente compresi, specialmente nel caso di modelli stocastici.

RIFERIMENTI BIBLIOGRAFICI

- [1] AGUIAR, M., KOOL, B., and STOLLENWERK, N. *Epidemiology of Dengue Fever: A Model with Temporary Cross-Immunity and Possible Secondary Infection Shows Bifurcations and Chaotic Behaviour in Wide Parameter Regions*. *Math. Model. Nat. Phenom.* **3** (2008), 48-70.
- [2] AJELLI, M., GONÇALVES, B., BALCAN, *et al.* *Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models*. *BMC Infect. Dis.* **10** (2010), 190.
- [3] AJELLI, M., MERLER, S., PUGLIESE, A., and RIZZO, C. *Model predictions and evaluation of possible control strategies for the 2009 H1N1v influenza pandemic in Italy*. *Epidemiol. Infect.* **139** (2011), 68-79.
- [4] BALL, F., MOLLISON, D., and SCALIA-TOMBA, G. *Epidemics with two levels of mixing*. *Ann. Appl. Probab.* **7** (2001), 46-89.
- [5] BALLARD, P. G., BEAN, N. G., and ROSS, J. V. *The probability of epidemic fade-out is non-monotonic in transmission rate for the Markovian SIR model with demography*. *Journal of Theoretical Biology* **393** (2016), 170-178.
- [6] DIEKMANN, O., HEESTERBEEK, J. A. P., and METZ, J. A. J. *On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations*. *J. Math. Biol.* **28** (1990), 365-382.
- [7] FERGUSON, N. M., CUMMINGS, D. a. T., CAUCHEMEZ, S., *et al.* *Strategies for containing an emerging influenza pandemic in Southeast Asia*. *Nature* **437**, 7056 (2005), 209-214.
- [8] GROPPI, M., and DELLA MARCA, R. *Modelli epidemiologici e vaccinazioni: da Bernoulli a oggi*. *Matematica, Cultura e Società*. *Rivista dell'Unione Matematica Italiana* **3**, **1** (2018), 45-59.
- [9] HELLEWELL, J., ABBOTT, S., GIMMA, A., *et al.* *Feasibility of controlling 2019-nCoV outbreaks by isolation of cases and contacts*. *medRxiv* (Jan 2020), <https://www.medrxiv.org/content/10.1101/2020.02.08.20021162v1>.
- [10] HETHCOTE, H. W. *Qualitative analysis of communicable disease models*. *Math. Biosci.* **28** (1976), 335-356.
- [11] IANNELLI, M., and MILNER, F. *The Basic Approach to Age-structured Population Dynamics. Models, Methods and Numerics*. Springer, 2017.
- [12] KERMACK, W. O., and MCKENDRICK, A. G. *A contributions to the mathematical theory of epidemics*. *Proc. R. Soc. London A* **115** (1927), 700-721. Reprinted in *Bull. Math. Biol.* **53** (1991), 33-55.
- [13] LI, Q., GUAN, X., WU, P., *et al.* *Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus Infected Pneumonia*. *New England Journal of Medicine* (Jan 2020), <https://www.nejm.org/doi/full/10.1056/NEJMoa2001316>.
- [14] MANFREDI, P., and D'ONOFRIO, A., *Eds.* *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*. Springer-Verlag New York, 2013.
- [15] PELLIS, L., BALL, F., and TRAPMAN, P. *Reproduction numbers for epidemic models with households and other social structures. I. Definition and calculation of R_0* . *Math. Biosci.* **235**, **1** (2012), 85-97.
- [16] RASMUSSEN, D. A., VOLZ, E. M., and KOELLE, K. *Phylodynamic Inference for Structured Epidemiological Models*. *PLOS Computational Biology* **10**, **4**, e1003570.
- [17] SCHWARTZ, I. B., and SMITH, H. L. *Infinite subharmonic bifurcations in an SEIR epidemic model*. *J. Math. Biol.* **18** (1983), 233-253.
- [18] WU, J. T., LEUNG, K., and LEUNG, G. M. *Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study*. *The Lancet* **395**, 10225 (Feb 2020), 689-697, [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).
- [19] WU, Z., and MCGOOGAN, J. M. *Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention*. *JAMA* (Feb 2020), <https://jamanetwork.com/journals/jama/fullarticle/2762130>.
- [20] CEREDA D., TIRANI M., ROVIDA F. *et al.* *The early phase of the COVID-19 outbreak in Lombardy, Italy*, [arXiv:2003.09320](https://arxiv.org/abs/2003.09320) <https://arxiv.org/abs/2003.09320>.
- [21] GANYANI, T., KREMER, C., CHEN, D., TORNERI, A., FAES, C., WALLINGA, J., HENS, N., *Estimating the generation interval for COVID-19 based on symptom onset data*. *medRxiv* [Preprint] <https://doi.org/10.1101/2020.03.05.20031815>.
- [22] LINTON, N.M., KOBAYASHI, T., YANG, Y., HAYASHI, K., AKHMETZHANOV, A.R., JUNG, S., YUAN, B., KINOSHITA, R., NISHIURA, H.. *Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data*. *Journal of Clinical Medicine*, **9**(2):538 (2020).



Andrea Pugliese

Nato nel 1956, si è laureato in Matematica all'Università di Roma La Sapienza nel 1978 e ha conseguito il dottorato di ricerca in Ecologia ed Evoluzione presso la State University of New York at Stony Brook nel 1986.

Dal 1983 è in servizio presso il Dipartimento di Matematica dell'Università di Trento, inizialmente come ricercatore, poi come professore associato e professore ordinario. Ha passato periodi di studio alla Cornell University (1989), come Fulbright Fellow alla Princeton University (1997) e come visiting professor alla Arizona State University (2016).

Fa parte del comitato editoriale del Journal of Mathematical Biology e del Journal of Biological Dynamics; è stato presidente della European Society for Mathematical and Theoretical Biology dal 2011 al 2014.

È autore o coautore di oltre 80 articoli su rivista (elenco su <https://orcid.org/0000-0002-3512-8560>) e di un libro di testo sulla matematica della dinamica di popolazioni.. È stato plenary o invited speaker a numerosi congressi internazionali nel campo della biomatematica.



Arcobaleno ad Antrodoto (Rieti). Licenza: Creative Commons Attribution 2.0 Generic.