

L'arrivo del coronavirus ha prodotto un'ondata di numeri e statistiche quotidiane. Mai come in questo momento i dati hanno giocato un ruolo tanto centrale nelle nostre vite. Gli italiani seguono con trepidazione il bollettino della protezione civile: numeri che parlano alla nostra speranza, che ci raccontano di sofferenze e lutti, e che forse possono fornirci una guida per il futuro. Il risultato è un pullulare di previsioni di picchi, e contrastanti previsioni su come – e quando -- sarà la nostra vita dopo l'emergenza, usando modelli statistici ed epidemiologici a volte contrastanti e non sempre giustificati.

Noi, un gruppo di fisici, informatici, data scientists e statistici degli istituti di ricerca triestini, abbiamo collaborato con la regione Friuli - Venezia Giulia all'analisi dei dati regionali aggregati sull'epidemia di Covid19. Questa esperienza ci ha permesso di toccare con mano cosa si possa leggere nei dati e quali siano le limitazioni delle varie tecniche alla base delle previsioni di cui si sente ogni giorno parlare.

I dati sui nuovi positivi sono fortemente influenzati da molti fattori, tra cui due fondamentali: il numero giornaliero di tamponi fatti (e come vengono fatti) e la distribuzione geografica della popolazione. La strategia (e il numero) di tamponi varia fortemente tra province diverse e in giorni diversi. Ciò non permette un'interpretazione diretta del numero dei positivi: bisogna tenere conto dei cambiamenti nelle modalità di rilevamento. Al contrario, dati aggregati a livello regionale, e, forse, più rilevanti e di qualità maggiore come i numeri di ricoverati (dato che si sta stabilizzando), o la frazione di positivi ricoverati su tutti i positivi (in costante discesa da una settimana) danno un'immagine più chiara della situazione. La struttura sociale della popolazione è ugualmente importante: particolarmente in regioni piccole: una parte rilevante degli incrementi proviene da situazioni come le case di riposo, ambienti dove il virus si diffonde facilmente, con effetti purtroppo devastanti ma non rappresentativi della circolazione del virus nella popolazione generale. Le analisi che semplicemente estrapolano previsioni dai numeri dei positivi danno risultati apparentemente chiari ma che rischiano di essere fortemente fuorvianti, soprattutto per previsioni a lungo termine, perché ignorano l'incertezza e l'incompletezza inerente in questi dati.

Un uso consapevole dei dati può comunque consentire un monitoraggio della situazione giorno per giorno e iniziare a permetterci di capire come sarà possibile tornare a una vita sociale. Si parla spesso di soluzioni "stop and go", o di test a tappeto, o ancora di tracciamento elettronico tramite cellulari. L'effetto di ciascuna di queste azioni è incerto: si deve ricorrere a modelli matematici per prevederlo. Ma i modelli contengono necessariamente semplificazioni rispetto alla realtà, che possono limitarne l'utilità nel fare predizioni accurate: ad esempio, il modello stop and go degli epidemiologi inglesi, di cui si è parlato molto in questi giorni, ignora misure di tracciamento e isolamento preventivo. Altri lavori ignorano la distribuzione geografica dell'epidemia, concentrandosi esclusivamente sull'effetto dei test preventivi. La realtà come sempre è più complessa: per ripartire sarà necessario mettere in campo una combinazione di interventi, differenziati a seconda del contesto territoriale.

Certamente un ampliamento della capacità di testare e tracciare individui sarà necessario, ma ugualmente importante sarà lo sviluppo di modelli matematici che rispecchino fedelmente le varie realtà locali, e la collaborazione effettiva con le strutture sanitarie locali che possono valutare in modo diretto la situazione. Quindi, per poter ripartire con il minimo di rischi, sarà necessario un investimento adeguato nello sviluppo di modelli numerici e statistici con consapevolezza dei loro limiti, e una stretta coordinazione con le autorità sanitarie locali per poter implementare strategie complementari in modo coordinato ed efficace.

La pandemia di Covid19 ha colto di sorpresa tutti i paesi occidentali; i sacrifici che stiamo facendo adesso sono necessari per ritornare ad un livello in cui future infezioni potranno essere controllate. Non facciamoci trovare impreparati una seconda volta. Speriamo che le esperienze maturate in questa fase portino a una più stretta collaborazione tra autorità sanitarie e la comunità scientifica per sviluppare strategie di raccolta dati appropriate, supportate da una comunicazione informata, trasparente e responsabile di analisi e modelli.

Prof Luca Bortolussi, UNITS & Saarland University (Informatico e Data Scientist)  
Dr Giulio Caravagna, UNITS & Institute of Cancer Research London (Informatico)  
Dr Leonardo Egidi, UNITS (Statistico)  
Dr Jacopo Grilli, ICTP (Fisico)  
Prof. Alessandro Laio, SISSA (Fisico)  
Dr Matteo Marsili, ICTP (Fisico)  
Prof. Francesco Pauli, UNITS (Statistico)  
Prof Guido Sanguinetti, SISSA & University of Edinburgh (Fisico e Data Scientist)  
Prof Alessandro Silva SISSA (Fisico)  
Prof. Nicola Torelli, UNITS (Statistico)  
Prof Roberto Trotta, SISSA & Imperial College London (Fisico e Data Scientist)