**Title**

The project is about “*Limits on forecasting the trend, peak and the end of an  
expanding epidemic while the epidemic is still spreading*”.

**Outline**

Here there is the outline of this presentation. I’ll start showing the model used to fit the data, that is a variant of SIR model but with confinement. I’ll explain how I fitted the COVID-19 data, using a parametric Bayesian model, and the sampler algorithm to sample model’s parameters, i.e. the Gibbs algorithm. At last results obtained Italy, Spain and France relatively at the first and the second waves.

**SCIR: SIR model with confinement**

(Slide 3)

The SCIR model includes the usual states of a SIR model plus a class C for individuals sent to confinement that are susceptible, but not infected. I compartimenti sono (***vedere schema***) ed i parametri sono (***vedere schema***, approvvigionamento alimentare, assistenza sanitaria o polizia o chi viola il confinamento). In figura è rappresentato il diagramma del modello epidemico, dove sono riportate le equazioni che descrivono la dinamica. In particolare, gli individui suscettibili (S) possono entrare o uscire dallo stato di confinamento © o diventare infetti (I). Gli individui infetti possono guarire ® o morire (D).

The fit I to data on officially diagnosed cases, which are automatically quarantined: the underlying assumption is that the real, mostly undetected, number of infections is proportional to the diagnosed cases.

(Slide 4)

Leggere slide. (…) Il sistema di equazioni si riduce ad un sistema di eq. in I ed S. La prima è un eq. diff lineare ordinaria a coeff. cost, mentre la seconda ha coeff. non costanti perché dipende da S(t). La prima eq. si risolve separando le variabili, per poi sostituirne la sol nella seconda, risolta anch’essa con metodo della separazione delle variabili.

(Slide 5)

Risoluzione eq. per gli infetti. La soluzione dipende da , da cui dipende tutto il comportamento dell’epidemia. Early stages : the epidemic spreads when  (as is the case of COVID-19), and the larger , the faster it does.

An important epidemiological message follows from this simple fact: only if the confinement is strong enough (p and q are sufficiently different so that ) can the epidemic be controlled; otherwise, it spreads until eventually decaying due to the standard SIR mechanism—the exhaustion of susceptible individuals. When confinement sets in,  gets tamed, eventually dropping to the value .

(Slide 6)

**Fitting COVID-19 data**

The data span two different regimes: unconstrained propagation of the epidemic, with , and a lockdown phase with effective parameters for the transition to the confined state. Since separated data for the number of recovered and deaths was unreliable, these two compartments were merged and jointly fitted . Here it is used a Bayesian approach to fit the data, assuming that the numbers of infected and recovered + dead are log-normally distributed with unknown variance and mean given by the expression for  obtained from the model. At the very early stages of the epidemic (before any recovery or death event), the total number of confirmed cases grows as  independently of the chosen model. Analysing this initial growth for every country in the world, it appears that  everywhere (doubling times larger than one day are reported in all cases). Thus, it is used informative priors for β and  (because ) (uniform distributions from 0 to 1 days−1) and vague priors for the rates q and p (uniform distributions from 0 to 5 days−1, we assume that changes faster than 1/5 d are meaningless in any compartmental model). The results are consistent with this assumption.). Also, we use noninformative priors for the variances.

In figura è schematizzato il modello.

(Slide 7)

**Gibbs sampler algorithm**

Dato un set di variabili random , aventi delle osservazioni associate, il campionamento di Gibbs consiste nel generare dei campioni a partire dalla distribuzione a posteriori. Quest’ultima può essere scritta utilizzando la regola di Bayes:

(*Full conditional* of )

"Proportional to" in this case means that the denominator is not a function of {\displaystyle x\_{j}} and thus is the same for all values of {\displaystyle x\_{j}}.

The algorithm samples from the posterior probability distribution based on the formulae of the full conditionals. The algorithm consists in:

1. pick a vector of starting values for the random variables, using the prior distributions of the variables *;*
2. select and draw a sample for this variable from its full conditional by fixing the values of the other variables to , i.e. sample from ;
3. select and draw a sample from , using the updated

value of ;

⋮

1. select and draw a sample from , using all the previously updated values;
2. build the vector ;
3. build a sequence of vectors using the above sampling procedure;
4. stop after steps.

In generale l’algoritmo di Gibbs è una procedura iterativa che genera una Catena di Markov di campioni.

(Slide 8)

**Fit results: before peak (Italy)**

In figura ho riportato le tracce e le posterior di ciascun parametro del modello. La linea in rosso rappresenta la mediana della distribuzione, riportata in tabella 1 per ciascun parametro insieme alla deviazione standard. In tale tabella è anche riportato il R cappuccio, ossia il potential reduction factor che è una misura della convergenza della catena.

(Slide 9)

**Fit results: before peak (Italy)**

The results of fitting real-time data until 01/04/2020

**Data description**:

In these figures are showed the active cases and the daily variation of recovered plus death cases, for Spain and Italy. The Spain data has been used to obtain same data of Castro et al., so no smoothing was applied because they didn’t. Instead for Italy data it has been performed a 7-days moving average smoothing. This solution significantly improved the quality of forecasting.

**Goal of this project -> Conclusion**

Epidemic spread is characterized by exponentially growing dynamics, which are intrinsically unpredictable. The time at which the growth in the number of infected individuals halts and starts decreasing cannot be calculated with certainty before the turning point is actually attained; neither can the end of the epidemic after the turning point. A susceptible-infected-removed (SIR) model with confinement (SCIR) illustrates how lockdown measures inhibit infection spread only above a threshold that is shown later in the presentation. The existence of that threshold has major effects in predictability: A Bayesian fit to the COVID-19 pandemic in Spain shows that a slowdown in the number of newly infected individuals during the expansion phase allows one to infer neither the precise position of the maximum nor whether the measures taken will bring the propagation to the inhibition regime. There is a short horizon for reliable prediction, followed by a dispersion of the possible trajectories that grows extremely fast. The impossibility to predict in the midterm is not due to wrong or incomplete data, since it persists in error-free, synthetically produced datasets and does not necessarily improve using larger datasets.

The parameters of the model can be estimated within a relatively narrow range using data available from COVID-19 pandemic. Yet, unavoidable uncertainties in those parameters, which determine the time at which growth is halted or the overall duration of the pandemic, propagate to the predicted trajectories, preventing reliable prediction of the intermediate and late stages of epidemic spread. This is the main message of this article because it transcends the model of choice.

This model does an excellent job in reproducing past data, but instead of taking most likely parameters values to draw a prediction, we estimate compatible ranges of variations in the parameters.

The main conclusion is that the deterministic nature of SIR-like models is misleading if aimed at describing the actual course of any pandemic: prediction of the past is achieved through suitable fitting of data, and different functions may work, but prediction of the future in the midterm cannot be trusted.

**SCIR: SIR model with confinement**

Here is performed a fit of I to data on officially diagnosed cases, which are automatically quarantined. The underlying assumption is that the real, mostly undetected, number of infections is proportional to the diagnosed cases.

In a sufficiently large population, the number of infected individuals at the initial stages of the infection is well below the population size. If we assume I(t)/N≪1 (1), then we can neglect the nonlinear term in the equation for the number of susceptible individuals and solve the model analytically. Within this approximation the equations determining the dynamics are:

The first two equations imply an approximate relation consequence of assumption (1), so we can write and reformulate the equations as

That are two *1° Order Constant Coefficient Linear ODEs* and they can be solved with separation of variable method (resolution).

The solution of this system is:

, ,

,

Where:

,

,

is the effective basic reproduction number modulated by the confinement—R0 being its value at the beginning of the epidemic.

**Uncertainty on Peak Occurrence Fitting Prepeak Data.**

The results of fitting real-time data are summarized in [**Fig. 2**](https://www.pnas.org/content/117/42/26190#F2). [**Fig. 2*A***](https://www.pnas.org/content/117/42/26190#F2) illustrates the fit to our analytical solution for the aggregated data of all Italy’s regions. Symbols are reported data, and the solid line represents the median of the distribution. Interestingly, quantiles 2.5% and 97.5% provide almost opposite conclusions: Either the epidemic curve “flattens,” or it keeps growing exponentially, albeit at a different rate. This is a consequence of the inherent variability of the fitted parameters—as summarized by their posterior distributions—and the exponential character of the epidemic. Similar conclusions can be drawn by inspection of the number of new deaths and recovered cases, ΔD+ΔR.

The systematic bending of the curve ([**Fig. 2*A***](https://www.pnas.org/content/117/42/26190#F2)), due to confinement in the framework of our model, does not guarantee that the epidemic is under control—hence, this information alone can be misleading in interpreting the effects of the measures applied. To emphasize this conclusion, we compute the posterior distribution of the time when the peak of the epidemic occurs. Analytically,

which, of course, is only meaningful when the epidemic gets eventually controlled by the confinement measures (i.e., if ). With parameter values inferred from [**Fig. 2*A***](https://www.pnas.org/content/117/42/26190#F2), confinement measures succeed at inhibiting the epidemic —which is the effect sought— in 53% of cases, while in 47% of cases they inhibiting its expansion.  [**Fig. 2*B***](https://www.pnas.org/content/117/42/26190#F2) displays the distribution of the day in which the epidemic reaches the maximum, conditional on it actually occurring.