Closed form solution of the SIR model for the COVID-19 outbreak in Italy

Riccardo Giubilei

Abstract

The CODIV-19 outbreak in early 2020 generated a tremendous effort of epidemiologists and researchers to fit the experimental data with the solutions of the SIR model equations [1] or with more sophisticated models. In this paper we show that under same hypotheses, a closed form solution exists that reasonably fits the experimental data for Italy, and the results can be extended to any other area.

Introduction

The COVID-19 epidemic in 2020 has aroused the interest of epidemiologists and researchers with the aim to validate the most common models that describe the evolution of a epidemic and to provide a reliable forecast of the epidemic dynamics ([2][3][4][5]).

Since the beginning of the outbreak, first in China and Far East, then in Italy and Europe, a lot of information was available on the Web [6]. Normally data about total infected, actual infected, recovered, dead are reported daily ([7][8] for Italy) and allow to carry out analyses to verify the fit of same model with experimental data.

The following figure 1 refers to Italy [7] and reports the daily new infected, from the 20th of February to mid-May. Data are affected by statistical fluctuations and by systematic biases that make them difficult to be interpreted. It is noted, for instance, that there is a sinusoidal behavior of the curve, around a mean value, with a period 7 days, that can be due to test results accumulated for a few days and accounted the days after.
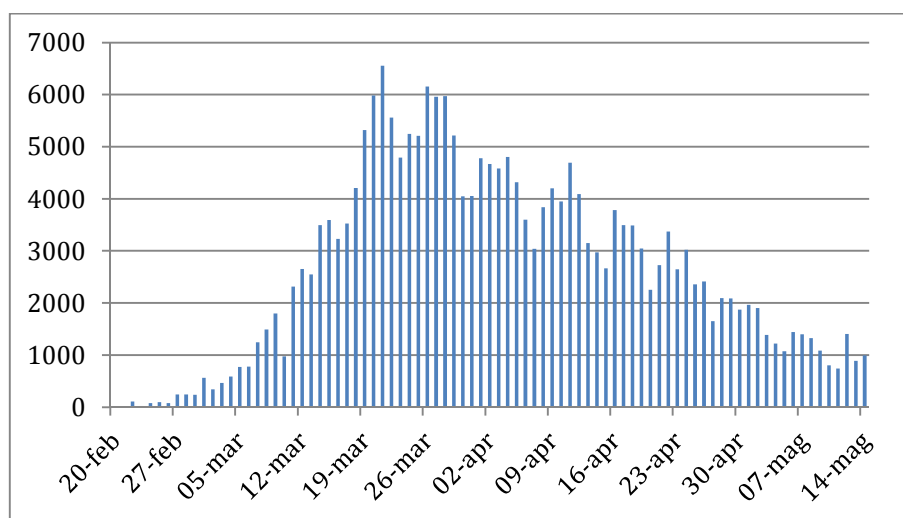

Figure 1: Daily new infected in Italy [7]

An interesting point is to find a curve that fits these experimental data. Among several functions considered, a very good fit has been obtained by using a Gumbel probability density function (pdf)[9], given by:

$$y(z) = G_0 e^{-(z+e^{-z})}$$

where $z = \frac{t-\mu}{b}$, being μ the mode of the function, that identifies its maximum. The question was to understand whether this "double exponential" function is simply a good fit of the experimental data or it is a good fit because this type of functions are the solution of the differential equations that govern the dynamics of the epidemic evolution.

The model taken for the investigation is a compartmental model called SIR model, proposed by Kermack and McKendrick in 1927 ([10],[11]). The model consists of three compartments: "S" for the number of susceptible, "I" for the number of infectious, and "R" for the number of removed (recovered or deceased) individuals.

The dynamics of the epidemic according to this model is a set of non-linear differential equations involving $S(t), I(t)$ and $R(t)$ and a closed form solution is not trivial ([12], [13]). In the following we make some assumptions (later on verified through the experimental data) and on the basis of these assumptions we find a very simple closed form solution, that fits the experimental data.

The SIR model

The SIR model can be expressed by the following set of ordinary differential equations [1]:

$$\frac{dS}{dt} = -\beta \frac{SI}{N} \qquad (1)$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I \qquad (2)$$

$$\frac{dR}{dt} = \gamma I \qquad (3)$$

where *S(t)* is the susceptible population, *I(t)* is the infected population and *R(t)* is the removed population (either by death or recovery), and *N = S(t)+ I(t) + R(t)*. N is constant, being the model without vital dynamics (birth and death).

Normally *S(t)* decreases and *R(t)* increases significantly, until the herd immunity is achieved. In the case of COVID-19 the social distancing has been put in place immediately and the epidemic slowed down in few weeks.

S(t), starting from an initial value S_0, is reduced by less than 0,06% worldwide and 0,4% in Italy [7], one of the most affected countries. On this ground we make the assumption that $S/N \cong 1$, therefore the second SIR equation becomes:

$$\frac{dI(t)}{dt} = \beta(t)I(t) - \gamma I(t) \qquad (4)$$

in which we assume that γ, the removal rate, is nearly constant and $\beta(t)$, the transmission rate, is a function of time, varying according to the social distancing measures put in place.

Equation (4) is a linear differential equation of the first order, whose solution, for $I(0) = I_0$ is ([14]):

$$I(t) = I_0 \, e^{-\int_0^t (\gamma - \beta(x))dx} \qquad (5)$$

Let us assume that $\beta(t)$ has an exponential shape, i.e. $\beta(t) = \beta_0 e^{-\alpha t}$. This conjecture will be verified by fitting experimental data. The exponent of the previous equation becomes:

$$-\int_0^t (\gamma - \beta(x))dx = \beta_0 \int_0^t e^{-\alpha x}\,dx - \gamma t = \frac{\beta_0}{\alpha}(1 - e^{-\alpha t}) - \gamma t$$

and finally:

$$I(t) = I_0 e^{\frac{\beta_0}{\alpha}} e^{-\gamma t - \frac{\beta_0}{\alpha} e^{-\alpha t}} \qquad (6)$$

The function is of the "double exponential" type and it looks like the Gumbel pdf, i.e. $G_0 e^{-(z+e^{-z})}$. We define now:

$$i_n(t) = \beta(t)I(t) \text{ , the new infected in the unit of time}$$

$$i_r(t) = \gamma I(t) \text{ , the removed ones, from the infected population, in the unit of time}$$

then:

$$\beta(t) = \frac{i_n(t)}{I(t)} \qquad (7)$$

$$\gamma = \frac{i_r(t)}{I(t)} \qquad (8)$$

The previous equations can be used to estimate $\beta(t)$ and γ from the experimental data.
We take the data relevant to Italy [7]. The t=0 corresponds to the 20th of February 2020. The unit of time is the day.
In the following diagram we show $\beta(t)$ calculated accordingly to (7), and based on the experimental data for $i_n(t)$, the new daily infected, and $I(t)$, the active cases, found in [7] .
Experimental data in fig. 2 confirms that $\beta(t)$ has a negative exponential behavior, thus validating the conjecture leading to equation (6).
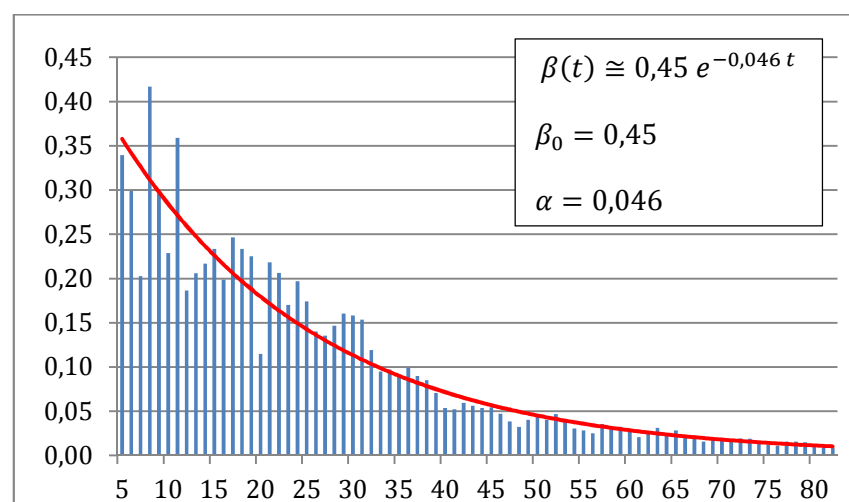

Figure 2: estimation of $\beta(t)$ – x-axis: number of days from the 20th of February

3

The γ parameter is not constant. Also averaging the daily value over 7 days, its moving average lies between 0,04 and 0,02 day^{-1}. We decided to retain a constant value, then $\gamma \cong 0,03\ day^{-1}$.
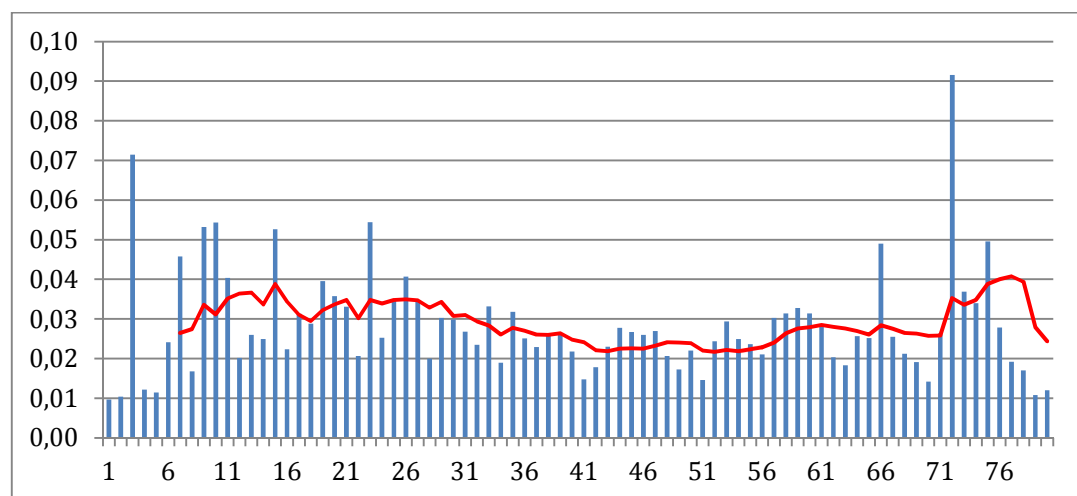

Figure 3: estimation of γ – x-axis: number of days from the 20th of February

Now β_0, α and γ are estimated and we can use equation (6) to fit the experimental data for $I(t)$, trying to tune the I_0 initial value. Figure 4 shows the results for I_0 = 65.
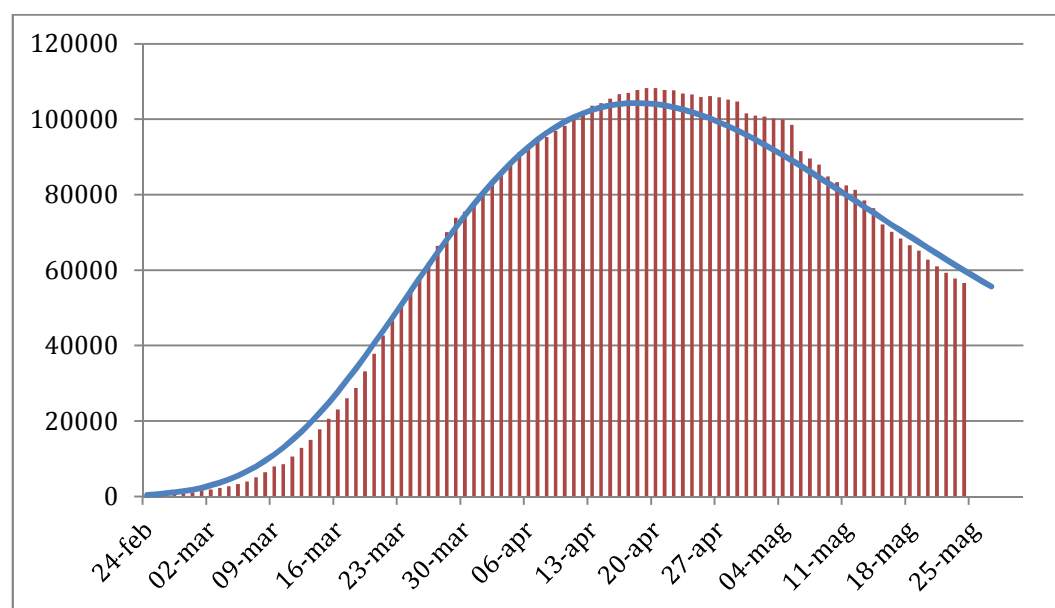

Figure 4: experimental [7] and theoretical curves for $I(t)$

By using the same parameters, the number of daily new cases has been calculated from the following equation:

$$i_n(t) = \beta(t)I(t) = I_0\beta_0 e^{\frac{\beta_0}{\alpha}} e^{-(\gamma+\alpha)t - \frac{\beta_0}{\alpha}e^{-\alpha t}} \qquad (9)$$

Also $i_n(t)$ is of the "double exponential" type. Next figure 5 shows an acceptable fit of the experimental data with equation (9).
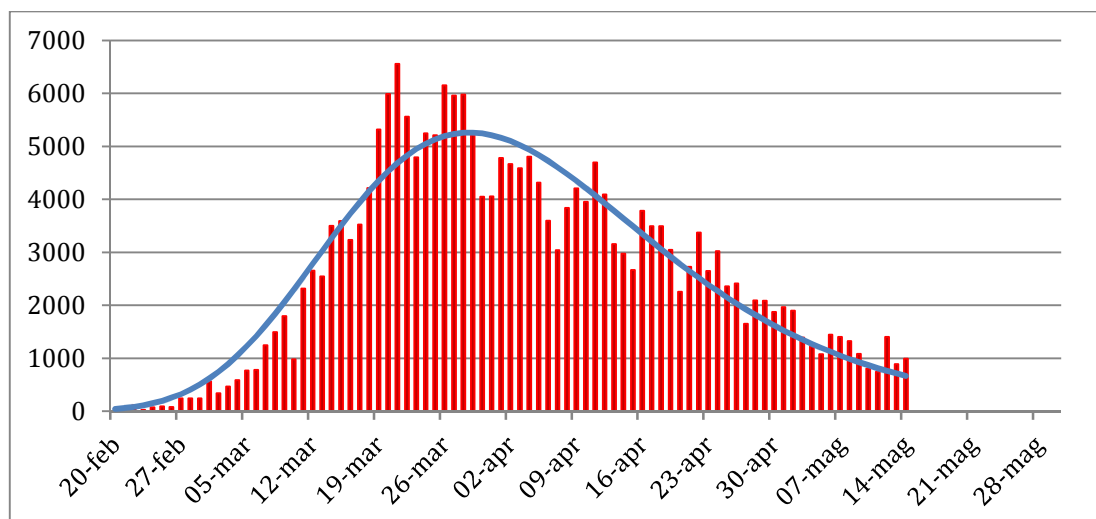
Figure5: experimental [7] and theoretical curves for $i_n(t)$, the daily new infected

Finally, by numerically integrating equation (9), we get the total number of infected during the epidemic, that is shown in figure 6:

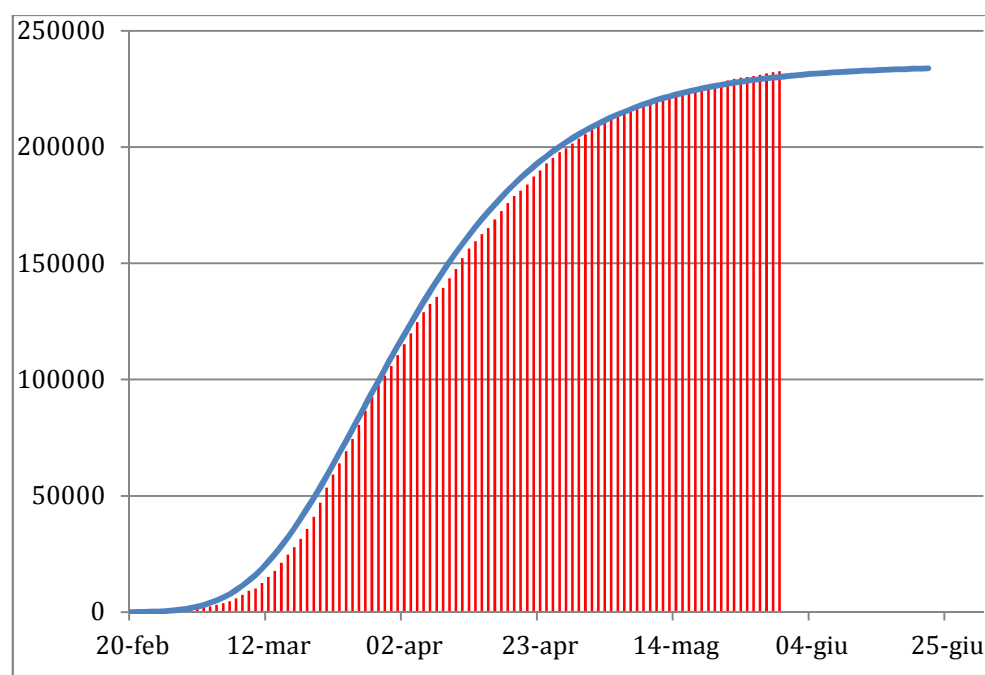$$I_{total}(t) = \int_0^t i_n(u)du \qquad (10)$$

Figure 6: experimental [7] and theoretical curves for $I_{total}(t)$, the total infected vs. time

The asymptotic value can be calculated by integrating equation (10) for $t \to \infty$, i.e.:

$$I_{total}(\infty) = \int_0^\infty i_n(u)du \qquad (11)$$

5

Putting (9) in (11) we get:

$$\int_0^\infty i_n(t)dt = I_0\beta_0 e^{\frac{\beta_0}{\alpha}} \int_0^\infty e^{-(\gamma+\alpha)t-\frac{\beta_0}{\alpha}e^{-\alpha t}} \, dt \qquad (12)$$

The integral is resolved making reference to the formulas (3.331) in [15] and it can be expressed in terms on the *Incomplete Gamma Function*, defined, according to Nielsen notation [16]:

$$P_x(u) \triangleq \int_0^x e^{-y}y^{u-1} \, dy$$

After same calculation we get:

$$I_{total}(\infty) = \int_0^\infty i_n(t)dt = I_0 e^{\frac{\beta_0}{\alpha}} \left(\frac{\beta_0}{\alpha}\right)^{-\frac{\gamma}{\alpha}} \cdot P_{\frac{\beta_0}{\alpha}}\left(1+\frac{\gamma}{\alpha}\right) \qquad (13)$$

All the parameters are known ($\beta_0 = 0,45, \ \alpha = 0,046, \gamma = 0,03, I_0 = 65$) and the asymptotic value given by (13) is around 234.000 total infected. It will be reached in the month of June in Italy.

Conclusions

On the basis of the classical SIR model, a closed form solution of the differential equations governing the epidemic dynamics has been found. The assumptions made to find the solution are $S/N \cong 1$, i.e. most of the population still lies on the susceptible compartment at the epidemic end, and $\beta(t)$ has a negative exponential shape. This last conjecture has been verified through the experimental data for Italy.

The solution is of the "double exponential" type and it looks like the Gumbel pdf, i.e. $G_0 e^{-(z+e^{-z})}$. The good fit with experimental data has been finally checked.

Bibliography

[1] https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology#The_SEIR_model

[2] Rajan, R.;"Predictions for COVID-19 outbreak in India using epidemiological models", medRxiv: https://doi.org/10.1101/2020.04.02.20051466, April 6, 2020

[3] Zhong, L. et al.: "Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China", IEEE Access, DOI: 10.1109/ACCESS.2020.2979599, March 24, 2020

[4] Waqas, M.; Farooq, M.;Ahmad, R.; Ahmad A.; "Analysis and Prediction of COVID-19 Pandemic in Pakistan using Time-dependent SIR Model", arXiv:2005.02353v1 [q-bio.PE] May, 5 2020

[5] Batista, M.; "Estimation of the final size of the coronavirus epidemic by the SIR model", https://www.researchgate.net/publication/339311383, February, 2020

[6] https://www.worldometers.info/coronavirus/

[7] https://datastudio.google.com/reporting/91350339-2c97-49b5-92b8965996530f00/page/RdlHB

[8] https://github.com/pcm-dpc/COVID-19

[9] https://en.wikipedia.org/wiki/Gumbel_distribution

[10] Kermack, W. O.; McKendrick, A. G. (1927). "A Contribution to the Mathematical Theory of Epidemics". Proceedings of the Royal Society A. 115 (772): 700-721.

[11] Hethcote, H. (2000). "The Mathematics of Infectious Diseases". SIAM Review. 42(4): 599–653, 2000

[12] Harko, T.; Lobo, F.; Mak, M. K. "Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates". Applied Mathematics and Computation. 236, 2014: 184–194

[13] Shabbir, G.; Khan, H.; Sadiq M. A.;"A note on Exact solution of SIR and SIS epidemic models" arXiv:1012.5035 , Dec. 2010

[14] Apostol, T. "Calculus", J. Wiley & Sons, New York, second edition, 1969

[15] Gradshtein, S. and Ryzhik, M. "Tables of Integrals, Series and Products", Academic Press, 1980

[16] Nielsen, N. "Handbuch der theorie der Gammafunktionen" (Teubner, Leipzig, 1906) (Chapter XV).