

An improved mathematical prediction of the time evolution of the Covid-19 Pandemic in Italy, with Monte Carlo simulations and error analyses

Ignazio Ciufolini¹ and Antonio Paolozzi²

1. Dipartimento di Ingegneria dell'Innovazione, University of Salento, Lecce and Centro Fermi, Rome, Italy
2. Scuola di Ingegneria Aerospaziale, Sapienza, University of Rome

Abstract

Here we present an improved mathematical analysis of the time evolution of the Covid-19 pandemic in Italy and a statistical error analyses of its evolution, including Monte Carlo simulations with a very large number of runs to evaluate the uncertainties in its evolution. A previous analysis was based on the assumption that the number of nasopharyngeal swabs would be constant, however the number of daily swabs has been increasing with an average factor of about five with respect to our previous analysis. Therefore, here we consider the time evolution of the ratio of the diagnosed positive cases to number of swabs, which is more representative of the evolution of the pandemic when the number of swabs is increasing or changing in time. We consider a number of possible distributions representing the evolution of the pandemic in Italy and we test their prediction capability over a period up to four weeks. The results show that a distribution of the type of Planck's black body radiation law provides very good forecasting. The use of different distributions provides an independent estimate of the uncertainty. We then consider five possible cases for the number of daily swabs and we then estimate the potential dates of a substantial reduction in the number of diagnosed positive cases. We then perform Monte Carlo simulations with 25000 runs to evaluate the uncertainty in the prediction of the date of a substantial reduction in the number of diagnosed daily cases. Finally, we present an alternative method to evaluate the uncertainty in our mathematical predictions based on the study of each region of Italy and we present an application of the Central Limit Theorem with 100000 runs to display the uncertainty in our mathematical predictions based on the analysis of each region.

1. Introduction

In a previous paper, we estimated the possible dates of a substantial reduction in the daily number of diagnosed positive cases of the Covid-19 pandemic based on the assumption that the number of nasopharyngeal swabs would remain roughly constant^{1,2}. At the time of our previous analysis (March 26), the average daily number of swabs from February 15 was about 9000 per day. However, from March 27 up to April 25, the average number of daily swabs was about 45000. Therefore, to study the evolution of the Covid-19 pandemic, we have to consider the analysis of the ratio of daily diagnosed positive cases per number of swabs. To possibly mathematically predict the evolution of the pandemic in Italy we can fit the ratio of cases per swab for several different distributions, however some distributions are more suitable than others for forecasting the future behavior of the pandemic. We consider the following distributions: Gauss, Beta, Gamma, Weibull, Lognormal and two of the type of the Planck's black body radiation law. The number of parameters chosen in some distributions is two to three. It turns out that Planck's law types distribution with three independent parameters provides the best predictions.

Furthermore, since the number of daily swabs depends on factors that are unknown to us, such as the daily availability of reagents and specialized personnel, we consider five possible cases for the daily number of swabs, in the previous version of this paper we also assume some possible time evolution in the number of daily swabs. We fit the time evolution of the positive cases per unit swab up to April 25, using the two best-fit-prediction distributions, i.e., Planck with three parameters and Gamma, along with the Gauss distribution. After analyzing the time evolution of the positive cases per unit swab using these three distributions and five conceivable number of daily swabs, we estimate the evolution in the number of diagnosed positive cases and the dates of a substantial reduction in such a daily number.

The different distributions that we use provide a possible independent way to estimate the uncertainty. Indeed, a basic problem is to mathematically estimate the uncertainty in the date of a substantial reduction of daily cases. For such a purpose, in section 4 we report the results of a Monte Carlo simulation with 25000 runs. Furthermore, in section 5 we present an alternative way to estimate the uncertainty in the dates of a substantial reduction in the number of daily cases which is based on the study of each region of Italy, where the conditions are quite different from each other, including the number of swabs per person. We finally report a way to display the analysis of the Italian cases using 100000 different combinations of the regions based on the Central Limit theorem.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

2. A mathematical analysis of the ratio of daily positive cases per swab

After analyzing the time trend of the ratio of daily positive cases to the number of daily swabs³⁻⁵, we found^{1,2} that this trend can be modeled by a Gauss distribution, however this time trend has also a certain amount of skewness that can be fitted by choosing a skewed distribution such as the Weibull, Log-normal, Beta and Gamma distributions, and also other distributions of the type of the Planck's black body law. This last one, which shows the best prediction capabilities, with three parameters a and b and c :

$$\frac{a t^b}{e^{c t} - 1}$$

where t is the time, is reported in Fig. 1. In Figs. 2a to 2e are reported the fits of the data with a distributions of the type of the Planck's black body law with two parameters (i.e., with the exponent of the time variable t equal to 3 as in the Planck's black body law), and with distributions of the type of the Gamma, Beta, Weibull, Lognormal respectively. The data can also be approximated by a function of the type of a Gauss function with three parameters a , b and c :

$$a e^{-\frac{(t-b)^2}{c}}$$

as shown in Fig. 2f.

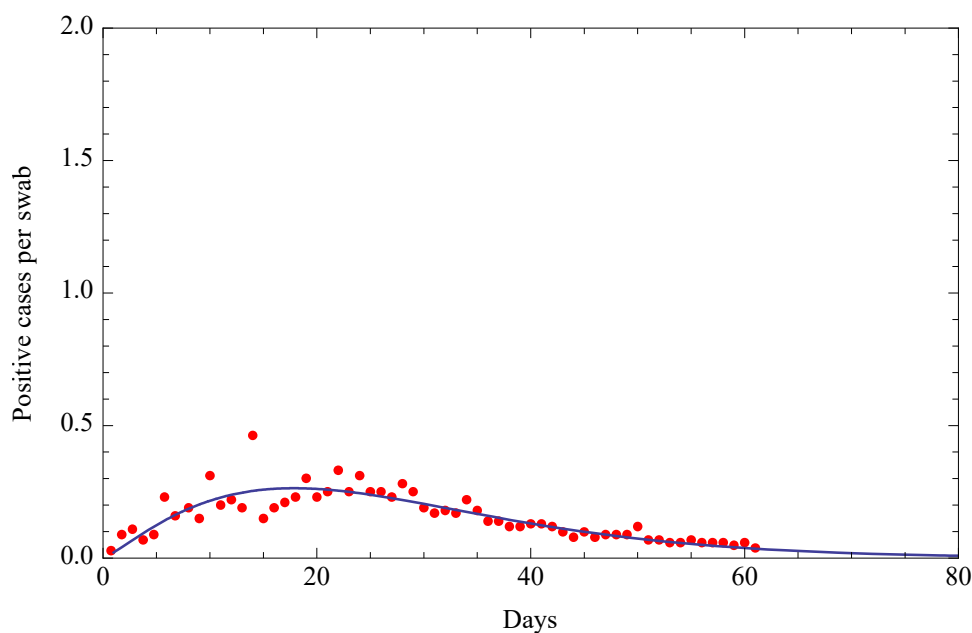


Fig. 1 Fit of the ratio of daily positive cases per swab obtained with a function of the type of the Planck's black body law with three parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.0446

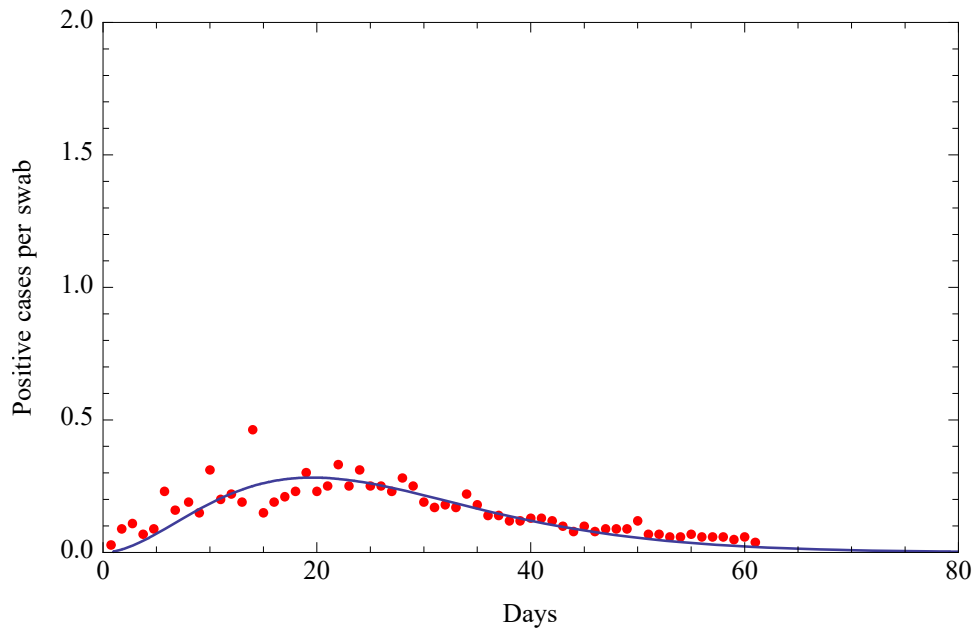


Fig. 2a Fit of the ratio of daily positive cases per swab obtained with a function of the type of the Planck's black body law with two parameters, with the exponent of the time variable t equal to 3 as in the Planck's law. The beginning date is February 25. Root Mean Square of the residuals is 0.050

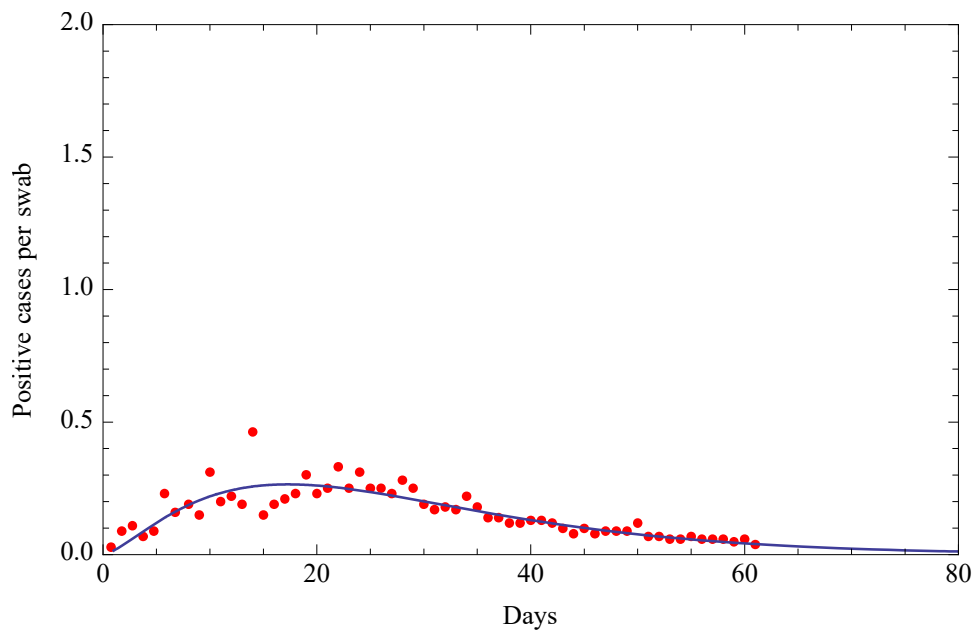


Fig. 2b Fit of the ratio of daily positive cases per swab obtained with a Gamma distribution with three parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.0450

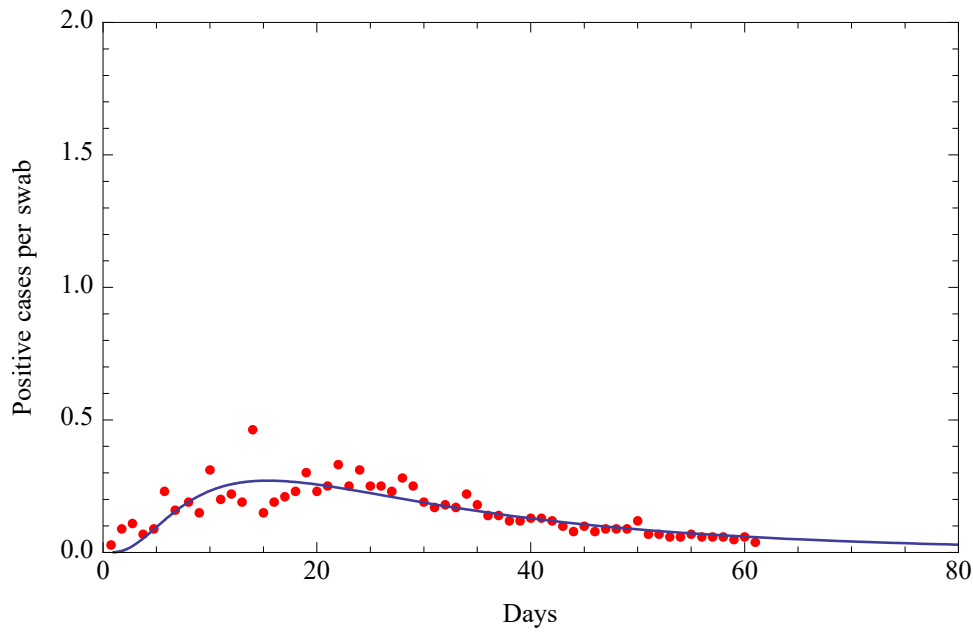


Fig. 2c Fit of the ratio of daily positive cases per swab obtained with a Beta distribution with two parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.046

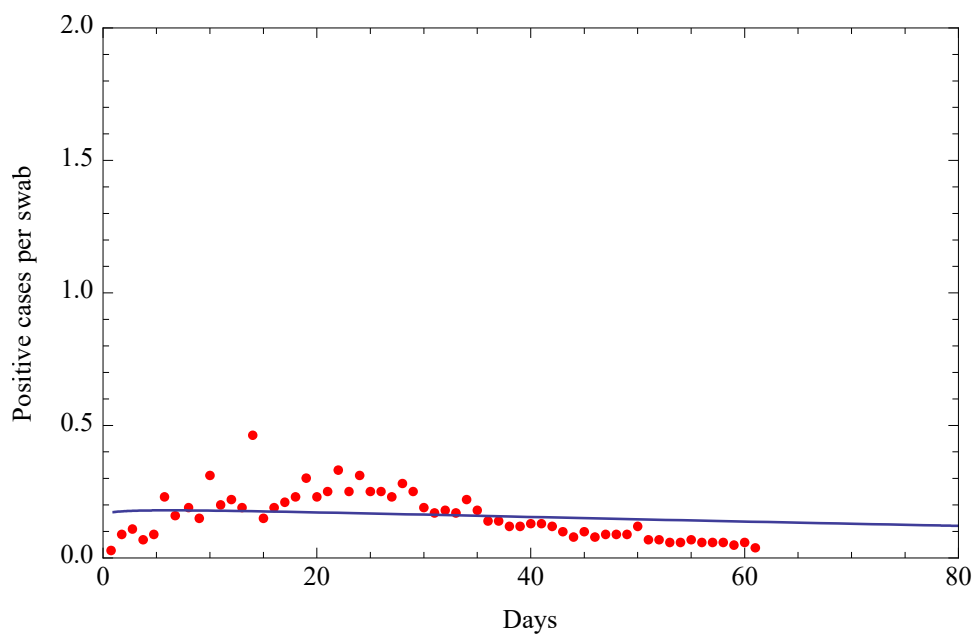


Fig. 2d Fit of the ratio of daily positive cases per swab obtained with a Weibull distribution with two parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.082

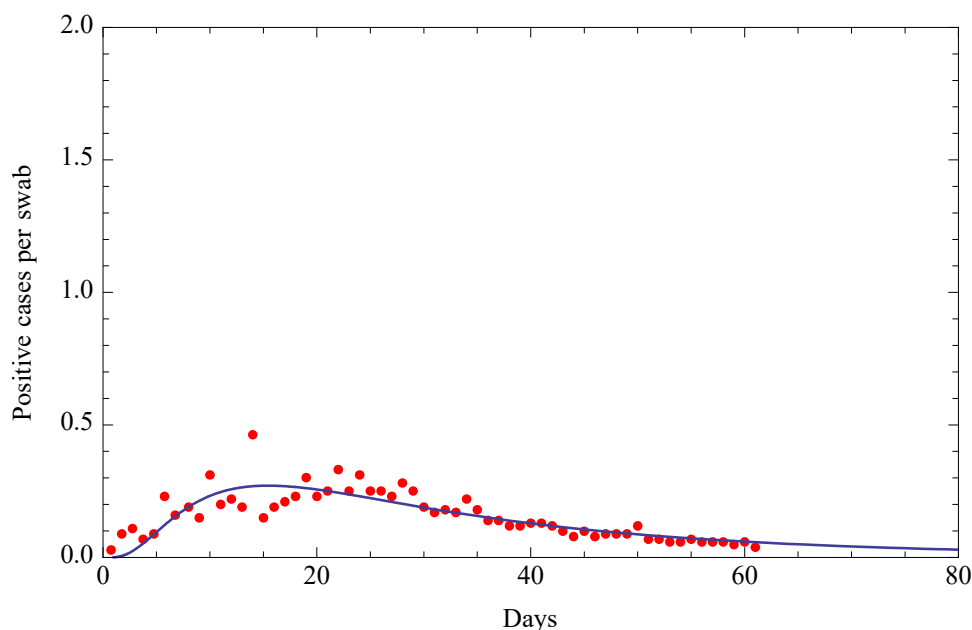


Fig. 2e Fit of the ratio of daily positive cases per unit swab obtained with a Lognormal distribution with three parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.048

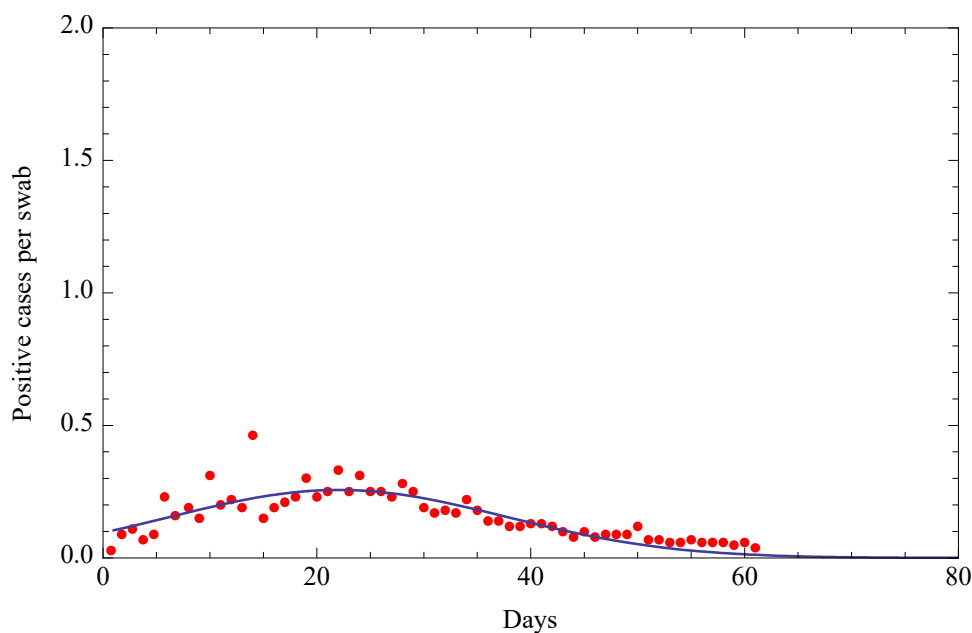


Fig. 2f Fit of the ratio of daily positive cases per swab obtained with function of the type of a Gauss function with three parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.049

2a Prediction capabilities of the distributions fitting the data

As a relevant test to select the distribution which is more suitable to predict the evolution in time of the daily positive cases per swab, we consider for each of the above distributions, three different intervals of daily ratios: from February 25 until April 18, from February 25 until April 11 and from February 25 until March 28, i.e., one week before the date of April 25, two weeks before April 25 and four weeks before April 25, respectively. We then test which one gives the best predictions for the measured ratios during the last five days (from April 21 until April 25) and the lowest absolute value of the mean and Root Mean Square (RMS) of the residuals (difference between the measured daily data and the fitting function) up to April 25, from April 18, April 11 and March 28, respectively. As shown in Tables 1 and 2, over a long period of about one month the distribution providing the best prediction is of the type of the Planck's black body law with three parameters. However, over a short period of up to two weeks the Gamma distribution shows a slightly better prediction behavior but less stable than the Planck's law with three parameters. The Beta and Plank with two parameters are also distributions with good prediction capability. The Weibull distribution does not fit well the data.

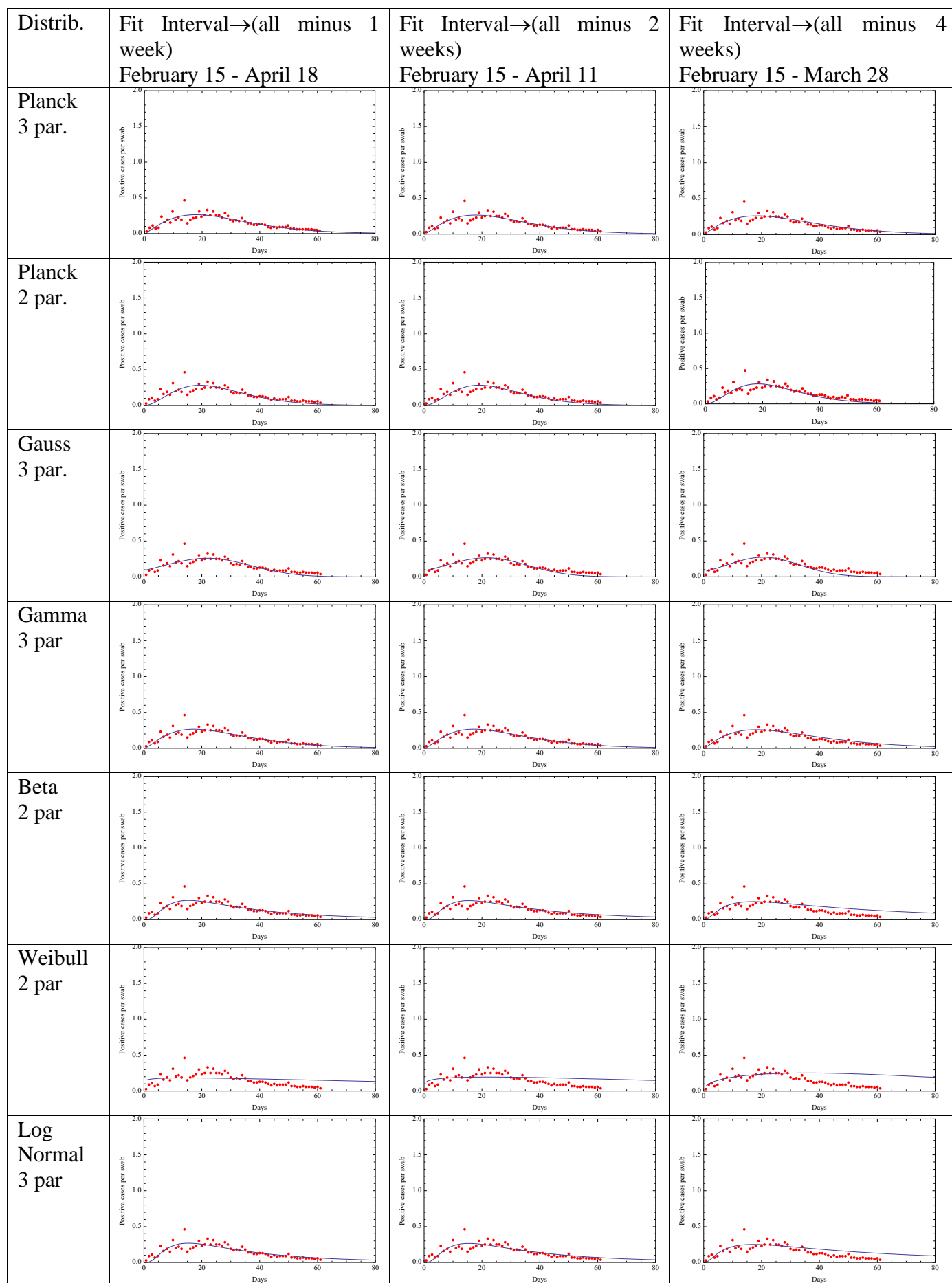


Table 1. Plots of the seven fitting distributions (rows) obtained over three different intervals of daily ratios (columns): from February 25 up to one week before the date of April 25, two weeks before April 25 and four weeks before April 25, respectively.

↓ Distribution	Fit up to April 25 minus the last 7 days. Mean and RMS between April 21 and April 25	Fit up to April 25 minus the last 14 days. Mean and RMS between April 21 and April 25	Fit up to April 25 minus the last 28 days. Mean and RMS between April 21 and April 25
Planck 2 par.	-0.022 0.0055	-0.024 0.0056	-0.029 0.0059
Planck 3 par.	-0.0061 0.0052	-0.0078 0.0052	0.0087 0.0050
Gauss 3 par.	-0.033 0.0057	-0.038 0.0061	-0.043 0.007
Gamma 3 par.	-0.0011 0.0052	-0.0016 0.0052	0.023 0.0050
Beta 2 par	0.0096 0.0050	0.011 0.0050	0.013 0.0050
Weibull 2 par	0.106 0.0067	0.124 0.0066	0.184 0.0060
Lognormal 3 par	0.0199 0.0055	0.026 0.0054	0.085 0.0054

Table 2. Mean and Root Mean Square of the residuals corresponding to the seven fitting distributions calculated over the last five days: April 21 to April 25.

↓ Distribution	Mean and RMS between April 18 and April 25	Mean and RMS between April 11 and April 25	Mean and RMS between March 28 and April 25
Planck 2 par.	-0.022 0.0046	-0.026 0.011	-0.026 0.015
Planck 3 par.	-0.0056 0.0044	-0.0073 0.011	0.018 0.016
Gauss 3 par.	-0.033 0.0047	-0.0398 0.011	-0.043 0.019
Gamma 3 par.	-0.00065 0.0044	-0.0014 0.011	0.029 0.016
Beta 2 par	0.011 0.0046	0.013 0.011	0.022 0.016
Weibull 2 par	0.10 0.0065	0.112 0.017	0.15 0.038
Lognormal 3 par	0.0195 0.0045	0.023 0.012	0.072 0.023

Table 3. Mean and Root Mean Square of the residuals corresponding to the seven fitting distributions calculated over the last week, last two weeks and last four weeks, respectively, i.e., up to April 25, from April 18, April 11 and March 28, respectively.

3. Modelling the daily swabs

Since the number of daily swabs depends on factors that are difficult to predict, such as the daily availability of reagents and specialized personnel, and since the actual number of positive cases is much higher than the measured ones⁶ and thus by increasing the number of daily swabs would also increase the number of diagnosed positive daily cases. we consider five possible cases for the daily number of swabs corresponding to some relevant situations. In a previous version^{1,2}, we

also tried to model the number of daily swabs using a Gaussian distribution, a distribution of the type of the Planck's law and a linear monotonic increasing distribution. The five cases we consider are: $\cong 9000$, i.e., the number of daily swabs equal to the mean of daily swabs between February 15 and March 26 (the date of our first analysis^{1,2}). $\cong 24000$, i.e., the mean between February 15 and April 25. $\cong 45000$, i.e., the mean between March 26 and April 25. $\cong 67000$, i.e., the maximum number of swabs up to April 25. 100000, i.e., an estimated possible upper bound to the number of daily swabs. In Table 4 are reported those five cases for distributions of the type of Planck's law and the Gamma one, i.e., the best distributions for what regards the prediction capability as reported in Tables 1-3, and for the Gauss distribution as a lower bound for the predicted date of a substantial reduction in the number of daily cases.

Daily swabs ↓	Distribution →	Normalized threshold for the number of daily cases	Day number and (date) with Gauss 3 parameters	Day number and (date) with Planck 3 parameters	Day number and (date) with Gamma 3 parameters
9000 mean up to 26 March		100 cases	62 (26 April)	77 (11 May)	81 (15 May)
24000 mean up to 25 April		267 cases			
45000 mean from 27 Mar to 25 Apr		500 cases			
67000 max value up to 25 April		744 cases			
100000 upper bound		1111 cases			

Table 4. Day when the number of positive cases becomes lower than 100 according to the prediction of the three distributions reported here. The beginning date is February 25 (included)

In our previous work^{1,2}, a substantial reduction of the daily cases was chosen to occur when the number of cases reduced to about 100. However, since the number of daily swabs is significantly changing and we may assume that the number of cases increases proportionally to the number of swabs, we have to normalize 100 by multiplying it by the ratio “number of swabs/9000”. The number 9000, as mentioned earlier, is approximately the average number of swabs up to March 26. Table 4 shows the dates of a substantial reduction of positive cases predicted with the three different distributions. The date of a substantial reduction of cases in the population should be independent on the number of swabs since it describes the real evolution of the pandemic at a certain stage. For this reason we introduce the normalization just described. In other words, the different numbers of daily cases reported in the second column of Table 4 represent the same actual condition of the pandemic independently on the number of daily swabs.

Since the number of daily swabs was increasing after March 26 by a factor of about five with respect to our previous analysis, a substantial decrease in the number of daily positive cases is reached later with respect to our previous mathematical prediction^{1,2}. Indeed, a better indication of the evolution of the pandemic is provided by the behavior of positive cases per swab as reported in Figs. 1 and 2. Since the maximum value of the number of positive cases per swab was 0.462 reached on March 9, we may then consider when such a number becomes 10 times smaller, i.e., lower than 0.0462. Indeed, on April 25, the ratio of positive cases per swab was 0.036, which is below its value of 0.081 reported two months ago, on February 26, and which shows that the pandemic is significantly reducing during the end of April.

In Fig. 3 we report a 3D representation of the number of daily positive cases as a function of the number of daily swabs, from 0 to 100000, and of the day, from February 25, using a distribution of the type of Planck's law.

In the next sections 4 and 5 the uncertainties associated with the mathematical predictions reported above are estimated.

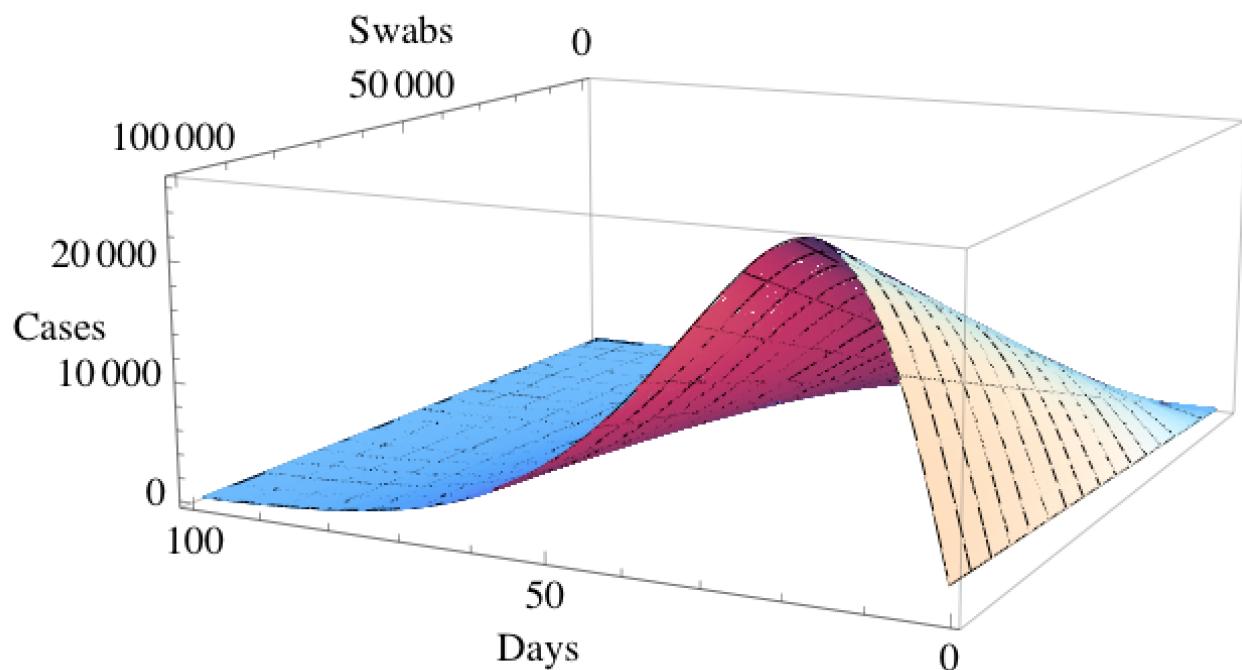


Fig 3. Representation in 3D of the number of daily cases as a function of the number of daily swabs (from 0 to 100000) and of the day, from February 25, for a distribution of the type of Planck's law with three parameters used to describe the positive cases per swab.

4. Monte Carlo simulations of cumulative positive cases of Covid-19 in Italy

Several uncertainties can influence the diagnosed cases of the Covid-19 pandemic, in addition to the number of nasopharyngeal swabs that have increased with time as explained in the previous section. To possibly estimate the uncertainties in the number of positive cases, we have used two methods: Monte Carlo simulations⁷⁻⁹, similarly to what done in previous works^{1,2} and a new method, described in the next section, using the study of each region of Italy. The uncertainty we consider in the Monte Carlo simulations is not the difference between the total number of the actual positive cases (which is unknown) and the diagnosed ones which can be one order of magnitude higher, or even more, than the actual cases. However, it is usual in statistics to use a sample as being representative of the population under study. The Monte Carlo simulations were performed for the number of positive cases.

For convenience to the reader, we summarize here the procedure used previously^{1,2}, the only difference being the number of simulations that have been largely increased from 150 to 25000. We have assumed a measurement uncertainty in the total number of positive cases equal to 20% of each daily number (Gaussian distributed). Then, a random matrix ($m \times n$) is generated, where n (columns) is the number of observed days and m (rows) is the number of random outcomes, which we have chosen to be 25000. Each number in the matrix is part of a Gaussian distribution with mean equal to 1 and sigma equal to 0.2 (i.e., 20% of 1), either row-wise and column-wise. So, starting from the n nominal values of the daily data, we generated n Gaussian distributions with 25000 outcomes, with means equal to the n nominal values and with 20% standard deviation. Then, for each of the 25000 simulations, those n values (corresponding to the cumulative positive cases, of n days) were fitted with a one parameter function of the type of the Gauss Error Function (see section 2) and we then determined the date of the flex with such fitted function for each simulation. Using the fitted function we also determined the date at which the number of daily positive cases will be less than a certain threshold that, for example, we have chosen to be 100 for the diagnosed positive cases. Finally, we calculated the standard deviation of the 25000 simulations. The value of the standard deviation is about one day.

In Fig. 5, is reported the histogram of the frequencies versus the day of a substantial reduction in the number of daily positive cases which has been chosen to be 100. The histogram approaches a Gaussian with mean equal to day $\cong 67$, approximately corresponding to what reported in Figs. 1 and 2 for a substantial reduction in the number of positive cases per unit swab. The standard deviation in Fig. 5 is approximately 1 day.

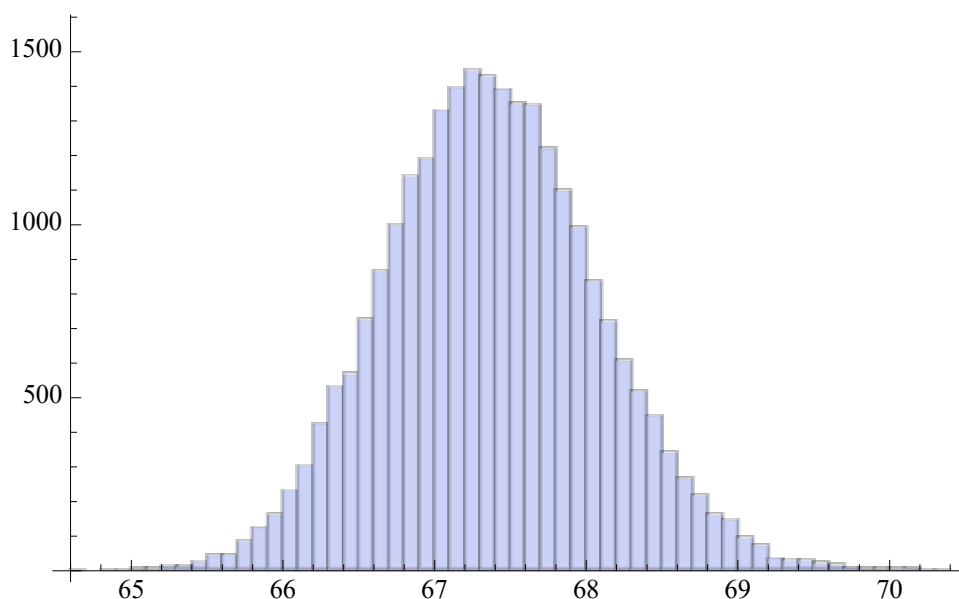


Fig. 5. Histogram of the 25000 runs of the Monte Carlo simulation for Italy: frequencies versus the day in which a substantial reduction in the number of daily cases occurs.

5. Analysis of uncertainty using each region of Italy

As a possible alternative approach to estimate the systematic uncertainty in the date of a substantial reduction in the positive cases, we independently analyze each of the twenty Italian regions from February 22, 2020 (included) until April 6, 2020 (included). Indeed, the number of daily swabs and other relevant conditions vary quite differently from one region to the other. We then evaluate the date of a reduction of the positive cases in each region below a certain threshold. The national threshold was chosen to be 100 cases, however, for each region, we normalized 100 for the number of positive cases in each region at the date of April 6 divided by the total number of national cases at the date of April 6 (i.e., 132547 cases). We then fitted the cumulative number of cases of each region using a function of the type of the Error Function including four parameters and finally, for each region, we obtained the date at which there is a substantial reduction of the diagnosed positive cases below the given threshold for each region. In Fig. 6 we report the 20 dates for each region. We then calculated the mean and the standard deviation of the 20 dates and we obtained a 1-sigma standard deviation of 9 days.

Day of reduction of cases below a threshold

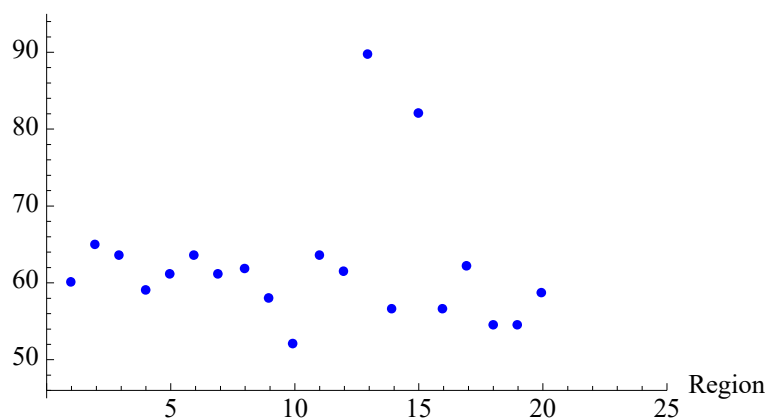


Fig. 6. Dates of a substantial reduction in the number of daily diagnosed positive cases for each of the 20 Italian regions. The 1-sigma standard deviation is about 9 days.

This result may also be displayed using the Central Limit Theorem. We have randomly chosen a sample of 30 regions, with repetition, out of the 20 regions, for 100000 times. We then calculated the mean of the day of a substantial reduction of the number of positive cases for each sample and we report the mean of each of the 100000 samples in the histogram of Fig. 7. We obtain a Gaussian with a standard deviation of 1.62 days, that, multiplied for the square root of 30, gives back a standard deviation of about 9 days as shown in Fig. 6. Since each region has a different number of daily swabs and quite different conditions, this figure might very roughly represent a kind of systematic error in the estimate of a substantial reduction in the number of daily positive cases in Italy. In this analysis we do not take into account the systematic increasing number of swabs since we just wish to evaluate the uncertainty in the date of a substantial reduction in the positive cases. This uncertainty seems to be consistent with the large variability in the date of a substantial reduction in the number of cases due to the assumed fitting distributions of the ratio of cases per swab and to the number of daily swabs, as reported in Figs. 1 and 2 and in Table 4.

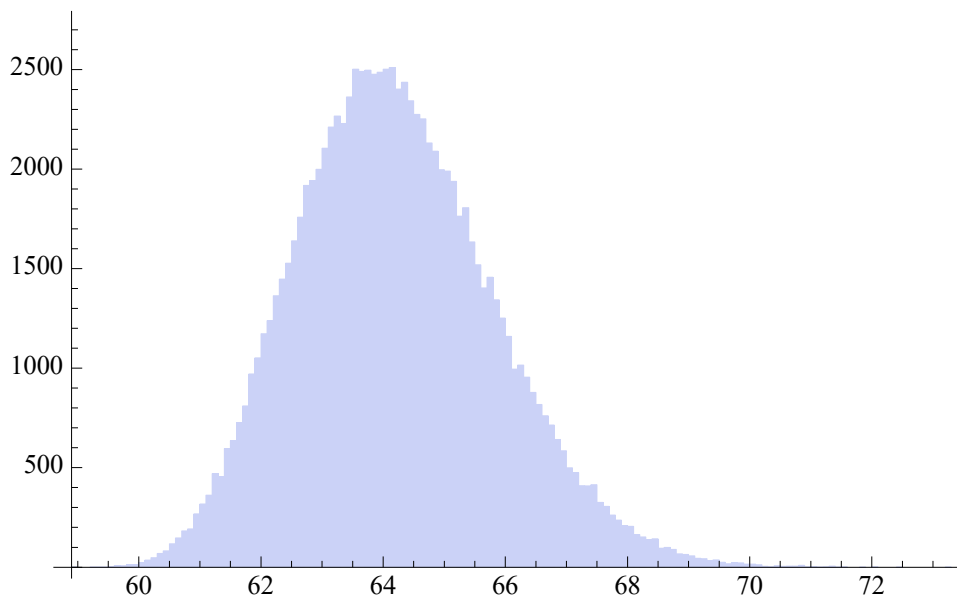


Fig. 7 Result of 100000 simulations using the 20 Italian regions for a possible estimate of the uncertainty in the date of a substantial reduction in the number of daily positive cases.

6. Conclusions

Since the number of daily swabs was in Italy highly increasing from March 26 to April 25, we fit the ratio of the positive cases to daily swabs using several functions, including the Gaussian, Weibull, Lognormal, Beta and Gamma distributions and a function of the type of the Planck's law, incidentally this last one fits well the number of daily positive cases in China. Considering that is practically impossible to predict of the evolution of the number of daily swabs, we consider five possible relevant cases for the number of daily swabs. By considering these five cases and the Gauss, Gamma and Planck-type distributions, the range of a substantial decrease in the number of daily positive cases goes from April 26 (in agreement with our previous findings) to May 15, depending on the chosen distribution. We have assumed, to characterize a substantial decrease of the pandemic, a threshold of 100 cases per day when an average of 9000 swabs per day are used. However, if for instance, the number of daily swabs is instead 67000 (the maximum number of swabs per day so far reached), the corresponding indication of that substantial decrease in the pandemic is given by a much higher number of cases, i.e., 744. Indeed, since the actual number of positive cases is much higher than the measured ones, by increasing the number of daily swabs, the number of diagnosed positive daily cases would also increase. To estimate the random uncertainty in the dates of a substantial reduction of the pandemic in Italy, we used a Monte Carlo simulation with 25000 runs which provides a random uncertainty of about one day. However, to possibly estimate some of the systematic uncertainties affecting our results, we also used the spread in a substantial reduction of the positive cases of each region below a certain threshold. Using this method, we found an uncertainty of about 9 days about the mean date, consistent with the estimates given above, by changing the fitting functions of the positive cases per swab.

Acknowledgments

We gratefully thank Richard Matzner (University of Texas at Austin) for helpful suggestions, Alessandro Paolozzi, and Claudio Paris (Centro Fermi).

References

1. Ciufolini, I., and Paolozzi A., Prediction of the time evolution of the Covid-19 Pandemic in Italy by a Gauss Error Function and Monte Carlo simulations. Submitted to BioRxiv on 03.26.2020 and transferred on 03.27.2020 to MedRxiv, doi: <https://doi.org/10.1101/2020.03.27.20045104>.
2. Ciufolini, I., and Paolozzi A., Prediction of the time evolution of the Covid-19 Pandemic in Italy by a Gauss Error Function and Monte Carlo simulations, *Eur Phys J Plus*, 2020; 135(4): 355. doi: [10.1140/epjp/s13360-020-00383-y](https://doi.org/10.1140/epjp/s13360-020-00383-y)
3. 1. <http://www.salute.gov.it/portale/home.html>
4. <https://www.worldometers.info/coronavirus/country/italy/>
5. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
6. Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang and Jeffrey Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2), *Science*, 16 Mar 2020, eabb3221
DOI: 10.1126/science.abb32214.
7. Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989). *Numerical recipes* (Vol. 3). Cambridge: Cambridge University Press.
8. Bevington, Philip R., and D. Keith Robinson. *Data reduction and error analysis for the physical sciences* (McGraw-Hill, New York, 1969).
9. Ciufolini, I., Monge, B. M., Paolozzi, A., König, R., Sindoni, G., Michalak, G., & Pavlis, E. C. (2013). Monte Carlo simulations of the LARES space experiment to test General Relativity and fundamental physics. *Classical and Quantum Gravity*, 30(23), 235009.