A decomposition of waves in time series of data related to Covid-19, applied to study the role of Alpha variant in the spread of infection

A decomposition of waves in time series of data related to Covid-19, applied to study the role of Alpha variant in the spread of infection

Amelia Carolina Sparavigna
Department of Applied Science and Technology, Politecnico di Torino

Here it is proposed a decomposition in components of the "waves" which appear in the time series of data related to Covid-19 pandemic. The decomposition is based on functions of κ-statistics; in particular the κ-Weibull is used. Fitted data are those of the "waves" ranging from August 2020 to April 2021 in the United Kingdom and from September 2020 to June 2021 in Italy. For the United Kingdom, the time series of daily infection shows a wave composed by two peaks. Among the many factors involved in the spread of infection, it seems that, in driving the onset of the second peak, the main role was played by the emergence of Alpha variant of Sars-Cov-2. For what concerns data from Italy, if we decompose the analysed "wave" in three peaks, we can observed that for two of them the role of the Alpha variant was relevant too.

Torino, October 25, 2021

Keywords: Time-series, Decomposition, Weibull functions, kappa-Weibull functions, Threshold, Variants of Sars-CoV-2.

Introduction

It seems that no formal definition exists for the "wave" of a pandemic, however the term is associated to a rising number of Covid-19 cases, which is characterized by a specific peak and then by a decline. In Ref.1 it is told that "Public health scientists first began using this term [wave] to describe different peaks and valleys of infections during influenza outbreaks in the late 1800s and the 1918-1929 Spanish flu." The Reference is also stressing that each "wave" has a different feature and can impact different populations, even within the same country. From data such as those given by www.worldometers.info, we can see that the trend of Covid-19 infection is generally given, from the second wave, by the composition of two or more peaks. For instance, in Italy we had a first wave which was characterized by an isolated peak, but from September 2020, the time series of data related to pandemic were characterized by a composition of some peaks (more than two peaks). In this framework, we can try to analyse data to distinguish the peaks in each "wave" - or the wavelets in a wave train - in order to have an instrument able of determining the onset of a specific component. In this manner, having the onset-time it is possible to identify what caused the surge of further infections. Let us note that, sometimes, the term "deconvolution" is applied to a process of decomposing peaks that overlap with each other.

The aim of the discussion here proposed is that of applying the κ-statistics to evaluate the onset of the peaks in the "waves". In particular we will compare the results obtained from fitting the data given by www.worldometers.info with information coming from Our World in Data, www.ourworldindata.org, concerning the variants of the virus. In the following, the function used for analysis is that proposed in [2], in the framework of κ-statistics.

Method

In the proposed method we will use the κ-Weibull function will be used. The data from time-series will be analysed using a 7-day moving average. Then, as in [2], the statistics is described by:

$$f_\kappa(t)=\frac{\alpha\beta t^{\alpha-1}}{\sqrt{1+\kappa^2\beta^2 t^{2\alpha}}}\ \exp_\kappa(-\beta t^\alpha) \quad (1)$$

where the κ-exponential is defined in the following manner:

$$\exp_\kappa(u)=\left(\sqrt{1+\kappa^2 u^2}+\kappa u\right)^{1/\kappa} \quad (2)$$

Eq. 1 is describing the κ-Weibull. Parameters α,β are the shape and scale indexes of Weibull distribution, whereas κ is the index of κ-distribution, that is the distribution introduced by G. Kaniadakis in [3,4].

In [2], where the first wave of Covid-19 had been analysed for China, Italy, Germany, Spain and United Kingdom, we have seen that (1) is properly fitting the data of time-series. Now, let us consider the case of a time-series which is characterized by the composition of two peaks. Data (courtesy www.worldometers.info) are from the surge of the second wave in the United Kingdom and are shown in the Figure 1. The daily number of cases is divided by the total number of cases, observed in the considered time period (in days).

In (1), we can add also the threshold parameter *T*, which is a feature of Weibull functions. Therefore, we have:

$$f_k(t|\alpha,\beta,T)=\frac{\alpha\beta(t-T)^{\alpha-1}}{\sqrt{1+\kappa^2\beta^2(t-T)^{2\alpha}}}\ \exp_\kappa(-\beta(t-T)^\alpha) \quad (3)$$

If we consider the case of data given in the Figure 1, let us try to use two functions (3), to fit the time-series, in the following form:

$$f=f_1+f_2=\xi f_{\kappa_1}(t|\alpha_1,\beta_1,T_1)+(1-\xi)f_{\kappa_2}(t|\alpha_2,\beta_2,T_2) \quad (4)$$

Parameter ξ , the mixing parameter, ranging from zero to 1, is used to generalize the addition of peaks, as proposed for the Weibull distribution [5]. It is also a rough manner to consider the fact that the set of population, involved by pandemic, changed for sure during the considered time period.

United Kingdom

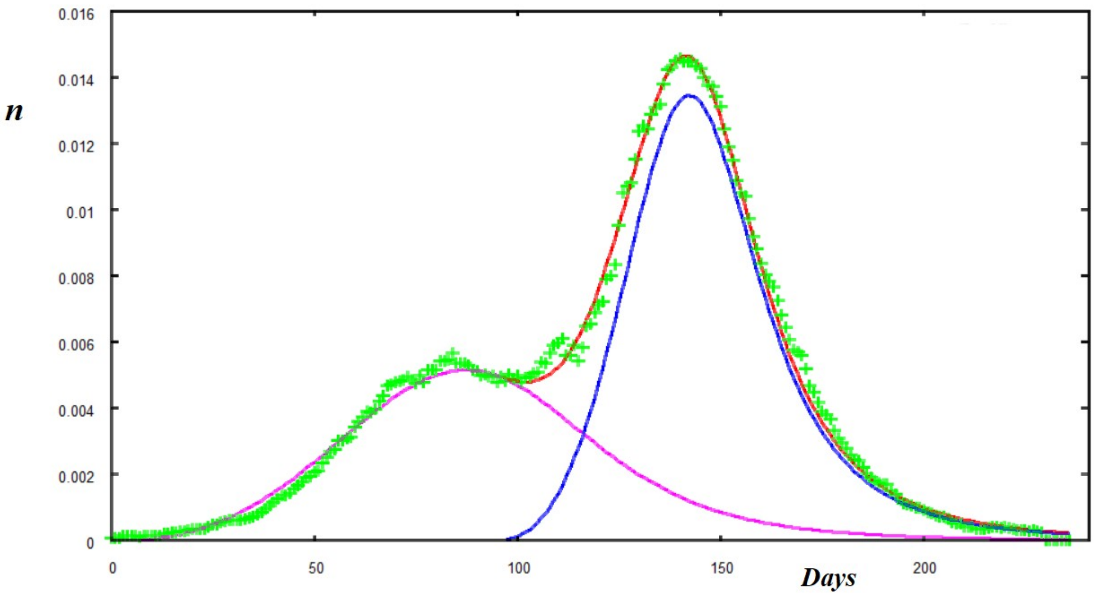

Figure 1 – Number of daily cases (victims) divided by the total number of cases. Parameters are $\kappa_1 = 0.52, \ \kappa_2 = 0.90$, $\alpha_1 = 3.50, \ \alpha_2 = 3.75$, $\beta_1 = 1.0 \times 10^{-7}, \ \beta_2 = 2.6 \times 10^{-7}$, $T_1 = 0, \ T_2 = 94$. $\xi = 0.4$.

Fitted data given in the Figure 1 are those ranging from August 2020 to April 2021 in the United Kingdom. Numbers are related to victims of the virus. Using the κ-Weibull we find the threshold time for the second peak at day 94, which is corresponding to the first week of December 2020.

In the Figure 2, the same approach has been applied to the number of the daily new infections. The data baseline has been shifted of 1180 cases. Time T_2 was November 21, 2020. This is the onset of the second peak in the Figure 2. A concomitant spread of the Alpha variant of Sars-CoV-2 can be observed in data given by Our World in Data. Among the many factors involved in the spread of infection, the Alpha variant of the virus seems the one which had the main role.

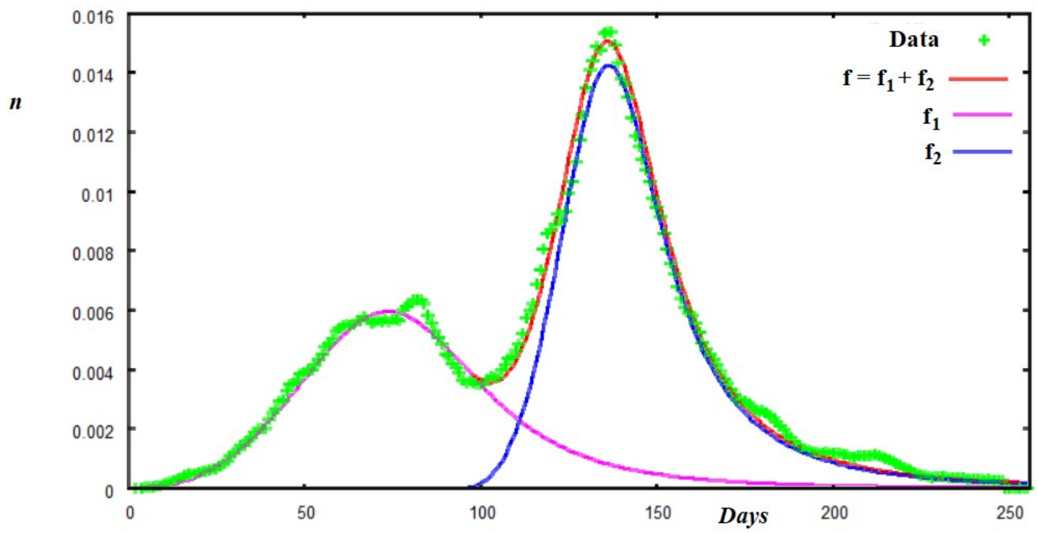

Figure 2 – Number of daily cases (infections) divided by the total number of cases, observed in the considered time period. Parameters are $\kappa_1=0.72, \kappa_2=1.25$, $\alpha_1=3.62, \alpha_2=3.85$, $\beta_1=1.0\times10^{-7}, \beta_2=2.6\times10^{-7}$, $T_1=0, T_2=94$. $\xi=0.4$. The first day was August 20, 2020. The day 94 was November 21, 2020.

As previously told, the decomposition is giving the onset of the second peak on November 21, 2020. From the web site Our World in Data, using this LINK we can evidence that, on 23 November 2020 for instance, the percentage of Alpha variant of Sars-CoV-2 was of 7.8 %. Using the data reported in the Table 1, we can argue that this variant had a leading role in the onset of the second largest peak of infections. Actually, the large increase in percentage of Alpha variant is the main feature of the second component of the wave train.

TABLE 1

1.96 % – November 9, 2020
7.8 % – November 23, 2020
----- T_2 -----
12.51 % – December 7, 2020
40.24 % – December 21, 2020
63.60 % – January 4, 2021
75.81 % – January 11, 2021
83.56 % – January 25, 2021
92.04 % – February 8, 2021
95.97 % – February 22, 2021

Table 1 – Percentage of Alpha Variant in analysed sequences for United Kingdom according to Our World in Data. Many, many thanks to this site and people involved in it.

Let us note that the web site, Our World in Data, tells that data are "The share of analysed sequences in the last two weeks that correspond to each variant group. This share may not reflect the complete breakdown of cases since only a fraction of all cases are sequenced".

Italy

Let us consider data concerning Italy from September 2020 to June 2021. In this case we need at least three peaks. In the following analysis, let us consider, just to propose the method, only three peaks. Therefore we use for the fit of data:

$$f=f_1+f_2+f_3=\xi_1 f_{\kappa_1}(t|\alpha_1,\beta_1,T_1)+\xi_2 f_{\kappa_2}(t|\alpha_2,\beta_2,T_2)+\xi_3 f_{\kappa_3}(t|\alpha_3,\beta_3,T_3) \quad (5)$$

In this case, $\xi_1+\xi_2+\xi_3=1$. In the Figure 3, a fitting is proposed based on three peaks. The data baseline has been shifted of 1237 cases. It is clear that to have a better result, more than three functions are required, such as a further adjustment of the baseline.

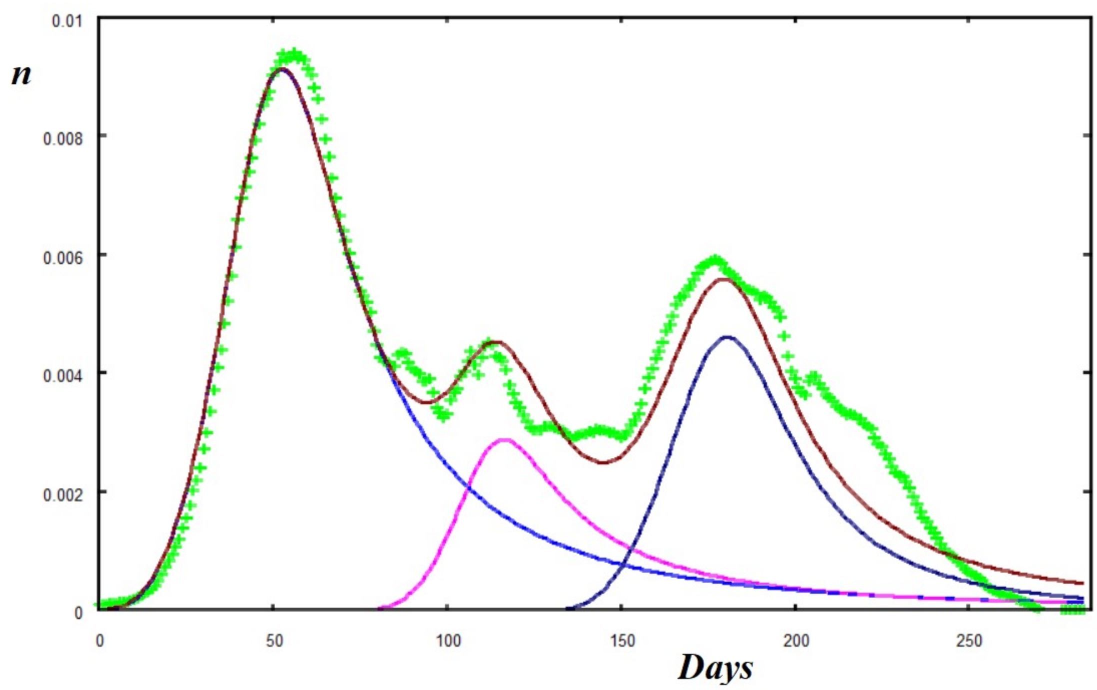

Figure 3 – Number of daily cases (infections) divided by the total number of cases, observed in the considered time period. Parameters are $\kappa_1=2.1$, $\kappa_2=3.2$, $\kappa_3=1.5$, $\alpha_1=4.00$, $\alpha_2=3.98$, $\alpha_3=3.80$, $\beta_1=0.55\times10^{-7}$, $\beta_2=1.10\times10^{-7}$, $\beta_3=1.50\times10^{-7}$, $T_1=0$, $T_2=75$, $T_3=130$. $\xi_1=0.57$, $\xi_2=0.18$, $\xi_3=0.25$. The first day was September 15, 2020. The day 75 was November 29, 2020. The day 130 was January 22, 2021.

Our World in Data tells that the percentage of Alpha variant in Italy was:

Table 2

0.77 % – October 12, 2020
1.78 % – October 26, 2020
0.51 % – November 9, 2020
1.16 % – November 23, 2020
----- T_2 -----
0.94 % – December 7, 2020
4.82 % – December 21, 2020
20.63 % – January 4, 2021
25.75 % – January 11, 2021
----- T_3 -----
25.34 % – January 25, 2021
59.81 % – February 8, 2021
44.74 % – February 22, 2021
69.38 % – March 8, 2021

If we use a decomposition of data concerning the pandemic wave from September 2020 to June 2021 in three components, comparing times T_2, T_3 with data giving the percentage of the Alpha variant in analysed sequences of Sars-Cov2 as in Table 2, we can argue that this variant had a relevant role in the second "wave" in Italy.

Here we have propose the use of κ-Weibull to decompose the peaks in the time series linked to Covid-19 pandemic. The work is preliminary. We have compared the onset-time of peaks with the percentage of Alpha variant in the analysed sequences. The use of more than three functions is required for the analysis of the pandemic in many cases. Also the mixing parameters could be refined, in the framework of κ-statistics.

References

[1] https://health-desk.org/articles/what-are-first-second-and-third-waves-of-infections
[2] Kaniadakis, G., Baldi, M.M., Deisboeck, T.S., Grisolia, G., Hristopulos, D.T., Scarfone, A.M., Sparavigna, A., Wada, T. and Lucia, U., 2020. The κ-statistics approach to epidemiology. Scientific Reports, 10(1), pp.1-14.
[3] Kaniadakis, G. (2002). Statistical mechanics in the context of special relativity. Physical review E, 66(5), 056125.
[4] Kaniadakis, G. (2001). Non-linear kinetics underlying generalized statistics. Physica A: Statistical mechanics and its applications, 296(3-4), 405-425.
[5] Razali, A. M., & Salih, A. A. (2009). Combining two Weibull distributions using a mixing parameter. European Journal of Scientific Research, 31(2), 296-305.