

The -statistics approach to epidemiology

*Original*

The -statistics approach to epidemiology / Kaniadakis, Giorgio; Baldi, Mauro M.; Deisboeck, Thomas S.; Grisolia, Giulia; Hristopulos, Dionissios T.; Scarfone, Antonio M.; Sparavigna, Amelia; Wada, Tatsuaki; Lucia, Umberto. - In: SCIENTIFIC REPORTS. - ISSN 2045-2322. - STAMPA. - 10:1(2020), pp. 1-14. [10.1038/s41598-020-76673-3]

*Availability:*

This version is available at: 11583/2853100 since: 2020-11-17T13:18:49Z

*Publisher:*

Nature Publishing Group

*Published*

DOI:10.1038/s41598-020-76673-3

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

Springer postprint/Author's Accepted Manuscript

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <http://dx.doi.org/10.1038/s41598-020-76673-3>

(Article begins on next page)



OPEN

# The $\kappa$ -statistics approach to epidemiology

Giorgio Kaniadakis<sup>1✉</sup>, Mauro M. Baldi<sup>2</sup>, Thomas S. Deisboeck<sup>3</sup>, Giulia Grisolia<sup>4</sup>, Dionissios T. Hristopoulos<sup>5</sup>, Antonio M. Scarfone<sup>6</sup>, Amelia Sparavigna<sup>1</sup>, Tatsuaki Wada<sup>7</sup> & Umberto Lucia<sup>4</sup>

A great variety of complex physical, natural and artificial systems are governed by statistical distributions, which often follow a standard exponential function in the bulk, while their tail obeys the Pareto power law. The recently introduced  $\kappa$ -statistics framework predicts distribution functions with this feature. A growing number of applications in different fields of investigation are beginning to prove the relevance and effectiveness of  $\kappa$ -statistics in fitting empirical data. In this paper, we use  $\kappa$ -statistics to formulate a statistical approach for epidemiological analysis. We validate the theoretical results by fitting the derived  $\kappa$ -Weibull distributions with data from the plague pandemic of 1417 in Florence as well as data from the COVID-19 pandemic in China over the entire cycle that concludes in April 16, 2020. As further validation of the proposed approach we present a more systematic analysis of COVID-19 data from countries such as Germany, Italy, Spain and United Kingdom, obtaining very good agreement between theoretical predictions and empirical observations. For these countries we also study the entire first cycle of the pandemic which extends until the end of July 2020. The fact that both the data of the Florence plague and those of the Covid-19 pandemic are successfully described by the same theoretical model, even though the two events are caused by different diseases and they are separated by more than 600 years, is evidence that the  $\kappa$ -Weibull model has universal features.

Many phenomena in a large variety of disciplines present apparent regularities described by well-known statistical distributions<sup>1</sup>. Some examples include the populations of cities, the frequencies of words in a text, the energy distribution in solids, the behaviour of financial markets, the statistical distribution of the kinetic energy in a gas, etc.<sup>2,3</sup>. Moreover, statistical approaches provide powerful tools for medical and epidemiological applications, since they allow predicting the behaviour of certain diseases<sup>4,5</sup>. Thus, such approaches are very useful for informing health policy and decision making, particularly regarding control and mitigation measures in response to the societal impacts of epidemics and pandemics.

While an epidemic is defined as “the occurrence in a community or region of cases of an illness clearly in excess of normal expectancy”<sup>6,7</sup>, a pandemic is defined as “an epidemic occurring over a very wide area, crossing international boundaries, and usually affecting a large number of people”<sup>6,7</sup>. Pandemics are large-scale outbreaks of infectious diseases which cause a growth in mortality over a wide geographic area.

In addition to the obvious health consequences, both epidemics and pandemics also cause economic stress and social hardship. It has been emphasized that the probability of occurrence of pandemics is increasing due to the high inter-connectedness of the modern world, which is facilitated by the ease of travel and a continuous increase of urbanisation<sup>6,8,9</sup>. Consequently, the international community is increasing its efforts toward the mitigation of the impacts of pandemics<sup>6</sup>. Some recent examples of pandemics are the 2003 SARS (Severe Acute Respiratory Syndrome), the 2014 West Africa Ebola epidemic, and the present COVID-19 (CORonaVirus Disease) pandemic caused by the SARS-CoV-2 virus.

<sup>1</sup>Dipartimento di Scienza Applicata e Tecnologia, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy. <sup>2</sup>Dipartimento di Informatica, Sistemistica e Comunicazione, Università di Milano-Bicocca, Viale Sarca, 336, 20126 Milan, Italy. <sup>3</sup>Department of Radiology, Harvard-MIT Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA 02129, USA. <sup>4</sup>Dipartimento di Energia “Galileo Ferraris”, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy. <sup>5</sup>School of Electrical and Computer Engineering, Technical University of Crete, Chania 73100, Greece. <sup>6</sup>Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche, c/o Dipartimento di Scienza Applicata e Tecnologia, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy. <sup>7</sup>Region of Electrical and Electronic Systems Engineering, Ibaraki University, Mito 316-8511, Japan. ✉email: giorgio.kaniadakis@polito.it

In this paper we formulate a statistical thermodynamics approach to epidemiology, which demonstrates the utility of  $\kappa$ -statistics for the analysis of epidemics and pandemics. First, we must consider certain facts which form the biomedical base of these phenomena. Below we summarise the principle causes of pandemics<sup>6</sup>:

- Pathogens such as influenza viruses that are capable of efficient transmission among humans, with a high potential to cause global and severe pandemics. They are characterized by a relatively long and asymptomatic infectious period, during which it is not possible to detect the infected persons and their movements.
- Pathogens such as Nipah virus which are capable of generating a moderate global threat. They are transmitted efficiently as a result of mutations and adaptation.
- Pathogens such as Ebola that are capable of causing local epidemics, with a non-negligible risk of evolving into a global pandemic.

Recently, pandemics have also found their origin in zoonotic transmissions of pathogens from animals to humans<sup>10</sup>.

The risk of pandemic spread is conditioned by the following factors<sup>11</sup>:

- pathogen factors such as genetic adaptation and mode of transmission;
- human factors such as population density, susceptibility to infection, travel, migration, poverty, malnutrition, and caloric deficits;
- factors related to public policy such as public health surveillance and measurements.

In order to better predict and mitigate the societal impact, the ability to quantify the morbidity and mortality associated with pandemics is very important. Consequently, statistical approaches based on available data that can lead to accurate models for predicting the behaviour of future pandemics are very useful. Historical data from past pandemics play a fundamental role, because they enable comparisons with theoretical models and can also be used to assess the performance of model-based forecasts. Historical records are often sparse and incomplete. Nonetheless, in all fields of research, scientists and engineers always search for “optimal” statistical distributions that can reliably predict the behaviour of different natural, engineered, and social systems<sup>12–14</sup>.

Since the beginning of the Covid-19 epidemic, various works in the scientific literature have presented statistical analyses of the epidemiological data. Most of these works focus on the number of infections<sup>15–21</sup> while the fatality curves have only been analyzed in a few studies. The analysis in the paper<sup>16</sup> focuses on the cumulative data (cumulative distribution function) of Covid-19, using the Richardson growth model which depends on three free parameters. The data related to Covid-19 deaths per day (i.e., the empirical probability density function) are analyzed in the paper<sup>15</sup> by adopting a four-parameter model which is based on non-extensive statistics. However, to our knowledge the development of a statistical model that admits explicit expressions for both the mortality cumulative distribution and the respective probability density function remains an open problem.

A second question concerns the mathematical features of suitable statistical models and in particular the behavior of the tails of the distribution functions. All of the existing models depend on three to five free parameters. However, a good fit of the data with a statistical model that depends on various free parameters is not sufficient proof of model validity. Two different model classes are mainly used to describe the Covid-19 data. The statistical distributions in the first class include the Weibull, Richardson, and SIR (Susceptible-Infected-Recovered) model functions. These depend on three to five free parameters and exhibit exponential or stretched exponential tails. The second class includes statistical distributions with two, three, or four free parameters that feature power-law tails. Models in this class include the Pareto, Zipf, Burr, Student and Tsallis distributions.

The present work aims to address the two problems that are described above, i.e., *analytical tractability* and *tail behavior*. We propose a statistical model for the analysis of the fatality curves that originated in physics but has also been successfully used in other fields such as econophysics, finance, and seismology. This model depends only on three free parameters but nonetheless provides excellent fits to the empirical curves obtained from the data. In addition, the model admits simple, closed-form analytic expressions for all the main statistical functions such as the probability density function, the cumulative distribution function, the survival function, the quantile function, the hazard function and the cumulative hazard function. Hence, this model overcomes the problem of analytical tractability and allows straightforward analysis of both daily and cumulative empirical data.

Regarding the second problem (tail behavior), the proposed  $\kappa$ -Weibull model is closely related to both classes mentioned above, but it does not strictly belong in either class. Indeed, the model's distribution function is a one-parameter continuous deformation of the Weibull distribution, which belongs in the first model class. However, the tail of the deformed distribution follows Pareto's power law as the distributions in the second class. Two of the three model parameters, namely  $\alpha$  and  $\beta$ , correspond to the standard Weibull parameters. The third parameter,  $\kappa$ , is related to deformation of the tail of the distribution into a power law;  $\kappa$  is directly linked to the Pareto exponent  $n$  of the survival function through the simple relationship  $\kappa = \alpha/n$ .

In order to test the validity of the  $\kappa$ -Weibull model we analyze various epidemics with different characteristics. First, we analyze the mortality data from the Florence plague that occurred in 1417<sup>22–24</sup>. These historical mortality data are of particular interest, because they were carefully recorded and allow testing the model with an epidemic process that evolved over several months. A second test of the  $\kappa$ -Weibull model involves the diffusion of Covid-19 in China<sup>25,26</sup>. The mortality curve derived from the Chinese Covid-19 data becomes flat after April 15, 2020. On the other hand, on April 17, 2020 China reported 1290 additional deaths which occurred during the entire cycle of the epidemic. However, no details were given regarding the temporal distribution of the fatalities. Hence, these data cannot be incorporated in our analysis. After April 15, 2020 only two cases of death have been reported to date (both in May 2020). However, these concerned infected individuals who had

returned to China from abroad. For this reason, it is considered that the mortality cycle from Covid-19 in China concluded on April 16. The Chinese data analyzed herein are the “partial data” until April 16, 2020 which exclude the additional 1290 deaths that were *a posteriori* reported on April 17.

To further test the  $\kappa$ -Weibull model, we present a systematic analysis of COVID-19 data from other countries such as Germany, Italy, Spain and United Kingdom<sup>25,26</sup>. The analyzed data from these countries spans the time window from the beginning of the epidemic in February 2020 until the beginning of August 2020. This temporal range completely captures the entire cycle of the first wave of the epidemic. Florence’s data are completely different from those of Covid-19 in terms of the epidemiological cause, the duration, the fatality rate and the historical time. The fact that the  $\kappa$ -Weibull model can successfully capture the behavior of data so different as those of Florence and Covid-19, as well as data from other fields of science (economics, finance, seismology) indicates that it possesses *universality features* which make it suitable for a wide range of applications.

## Methods

The fundamental difficulty faced by mathematical approaches to epidemiology is that all forecasts strongly depend on the model employed, the parameter estimates, as well as the choices of the initial conditions. There are both deterministic and stochastic epidemiological models<sup>5,27</sup>. The  $\kappa$ -deformed model that we propose herein has its roots in statistical mechanics and is therefore stochastic by construction.

**The  $\kappa$ -deformed exponential function.** Let us consider the function  $n(t)$  to represent the number density of deaths at any given time  $t \in \mathcal{D}$ , where  $\mathcal{D}$  is the temporal domain of interest. In most cases of interest  $\mathcal{D} = [t_0, \infty)$ , where  $t_0 \geq 0$ . The probability of death within a short time interval  $\delta t$  around  $t$  is given by  $f(t)\delta t$ , where  $f(t)$  is the *probability density function* (pdf). Then, the respective number density is given by  $n(t) = Nf(t)$ , where  $N$  is the total number of deaths.

In statistical mechanics a general rate equation for  $f(t)$  is the first-order linear ordinary differential equation (ODE)

$$\frac{df(t)}{dt} = -r(t)f(t), \quad (1)$$

where the function  $r(t)$  is the *decay rate*. The solution of the above ODE is the exponential

$$f(t) = c \exp \left( - \int_{t_0}^t r(t') dt' \right), \quad (2)$$

with the standard normalisation condition which determines the constant  $c$ :

$$\int_{t_0}^t f(t') dt' = 1. \quad (3)$$

The normalization also enforces a constraint on the number density, i.e.,  $N = \int_{t_0}^{\infty} n(t) dt$ .

In the context of the exponential solution, the following three simple cases must be considered.

- The Exponential Model: this model is fundamental in every branch of science; indeed, it allows us to describe a great variety of phenomena, from elasticity to electricity, from nuclear decay to thermal transient response, to name a few. The exponential model can be obtained for constant decay rate, i.e.,

$$r(t) = \beta, \quad (4)$$

which leads to the following exponential pdf:

$$f(t) = \beta \exp(-\beta t), \quad (5)$$

with  $\mathcal{D} = (0, +\infty)$ .

- Power-law Model (Pareto distribution): Zipf<sup>28–31</sup>, in his studies on the size distribution of cities, incomes and word frequencies, pointed out the notion of regularity in the distribution of sizes. In a great variety of these cases the distributions follow a power law with an exponent close to  $-1$ , also known as Zipf’s Law<sup>32–38</sup>. Such power law distributions have been considered with increasing interest in the description of regular distributions. Distributions with exponents different from  $-1$  are known as Pareto’s law<sup>3,36,39</sup>. In general, distributions with power-law tails are known as *heavy-tailed*, *fat-tailed* or *subexponential* distributions, in juxtaposition to distributions whose tails decay exponentially<sup>2</sup>.

Perline introduced the following classification of power laws<sup>40</sup>:

- A power law is *strong* if the distribution is a power law over the entire domain of definition;
- A power law is *weak* if only part of the distribution is fitted by a power law;
- A power law is *false* if only a highly truncated part of the distribution is approximated by a power law (in the scientific literature there are many examples of false power laws<sup>41</sup>).

Pareto obtained his *Type I model*, named the *Pareto law* by fitting the available data in his time on increasing social inequalities. He concluded that only economic growth can increase the income of the poor and decrease inequality<sup>42</sup>. Today, we know that his conclusions were partially right and partially wrong, because

economic growth and equity are strictly related to the social relations of production, to the technological level and institutional structures of the economy, and to the composition, accumulation and distribution of human capital, as well as the quality and accessibility of the educational and financial structures in place<sup>43</sup>. Moreover, since Pareto disseminated his approach in 1897<sup>44</sup>, the application of heavy-tailed distributions in economics has been developed<sup>45</sup>.

The Pareto pdf  $f(t)$  is obtained from Eq. (1) by inserting the following time-dependent decay rate function

$$r(t) = \frac{p}{t}, \quad p > 1, \quad (6)$$

which leads to a power-law solution for the pdf, i.e.,

$$f(t) = \frac{p-1}{t_0} \left( \frac{t_0}{t} \right)^p, \quad p > 1, \quad (7)$$

in the domain  $\mathcal{D} = (t_0, +\infty)$ , with  $t_0 > 0$ :

- $\kappa$ -Exponential Model: This model, which is based on a fundamental approach derived from relativity<sup>46</sup>, has proved useful in many applications. Experimental evidence suggests that probability density functions should resemble the exponential function for  $t \rightarrow 0$ . However, for  $t \rightarrow 0$  the Pareto pdf diverges. On the other hand, for high values of  $t$  many experimental results show a Pareto-like pdf with power law tails instead of exponential decay. Consequently, for  $t \rightarrow 0$  it follows that  $r(t) \sim \beta$  while for  $t \rightarrow +\infty$  it follows  $r(t) \sim p/t$ . So, the actual decay rate function  $r(t)$  should smoothly interpolate between these two regimes; a good proposal for  $r(t)$  has been introduced in the context of special relativity, where the function  $r(t)$  is given in terms of the Lorentz factor. We recall the expression of the Lorentz factor  $\gamma_\kappa(q) = \sqrt{1 + \kappa^2 q^2}$ ; this expression involves the dimensionless momentum  $q$  where the parameter  $\kappa$  is the reciprocal of the dimensionless light speed  $c$ , i.e.  $\kappa \propto 1/c$ . After posing  $r(t) = \beta/\gamma_\kappa(\beta t)$  or more explicitly

$$r(t) = \frac{\beta}{\sqrt{1 + \kappa^2 \beta^2 t^2}}, \quad (8)$$

it follows that for  $t \rightarrow 0$  the decay rate  $r(t)$  approaches the exponential regime, i.e.  $r(t) \sim \beta$ . On the other hand, for  $t \rightarrow +\infty$  it follows that  $r(t)$  approaches the decay rate of the Pareto model, i.e.  $r(t) \sim 1/\kappa t$ .

The solution of the rate equation in this case yields the following pdf

$$f(t) = (1 - \kappa^2) \beta \exp_\kappa(-\beta t), \quad (9)$$

where the  $\kappa$ -deformed exponential function is given by

$$\exp_\kappa(t) = \left( \sqrt{1 + \kappa^2 t^2} + \kappa t \right)^{1/\kappa}, \quad (10)$$

with  $0 < \kappa < 1$ . It is important to note that in the  $\kappa \rightarrow 0$  limit and in the  $t \rightarrow 0$  limit the function  $\exp_\kappa(t)$  approaches the ordinary exponential  $\exp(t)$ , i.e.

$$\exp_\kappa(t) \underset{\kappa \rightarrow 0}{\sim} \exp(t), \quad (11)$$

$$\exp_\kappa(t) \underset{t \rightarrow 0}{\sim} \exp(t). \quad (12)$$

On the other hand the function  $\exp_\kappa(-t)$  for  $t \rightarrow +\infty$  presents a power-law tail, i.e.

$$\exp_\kappa(-t) \underset{t \rightarrow +\infty}{\sim} (2\kappa t)^{-1/\kappa}. \quad (13)$$

Furthermore, the  $\kappa$ -exponential satisfies the following identity

$$\exp_\kappa(t) \exp_\kappa(-t) = 1, \quad (14)$$

in analogy with the standard, non-deformed, exponential.

The  $\kappa$ -exponential represents a very powerful tool which can be used to formulate a generalized statistical theory capable of treating systems described by distribution functions that exhibit power-law tails<sup>46–48</sup>. The mechanism generating the  $\kappa$ -exponential function is based on first principles from special relativity, and therefore the new function appears very promising for physical applications. Generalized statistical mechanics, based on the  $\kappa$ -exponential, preserves the main features of ordinary Boltzmann-Gibbs statistical mechanics which is based on the ordinary exponential through the Boltzmann factor. For this reason, it has attracted the interest of many researchers over the last two decades who have studied its foundations and mathematical aspects<sup>49–58</sup>, the underlying thermodynamics<sup>59,60</sup>, and specific applications of the theory to various fields. A non-exhaustive list of applications includes those in quantum statistics<sup>61–64</sup>, in quantum entanglement<sup>65,66</sup>, in plasma physics<sup>67–73</sup>, in nuclear fission<sup>74–77</sup>, in astrophysics<sup>78–81</sup>, in quantum gravity<sup>82–87</sup>, in geomechanics<sup>88,89</sup>, in genomics<sup>90,91</sup>, in complex networks<sup>92–94</sup>, in economy<sup>95–99</sup>, in finance<sup>100–103</sup>, as well as in reliability analysis and seismology<sup>104–106</sup>.

**The  $\kappa$ -deformed statistical model.** Given a pdf  $f(t)$  which represents the death rate, the *cumulative distribution function* (cdf)  $F(t) : \mathcal{D} \rightarrow [0, 1]$  represents the probability of death between the initial time and the

current time  $t$ .  $F(t)$ , which is also known as the *lifetime distribution*, is given by means of the following integral  $F(t) = \int_{t_0}^t f(t') dt'$ .

Conversely, the pdf  $f(t)$  is given by the derivative of  $F(t)$ . In the following, we will assume without loss of generality that  $t_0 = 0$ . The complement of  $F(t)$ , i.e.,  $S(t) = 1 - F(t)$  is known as the *survival function*, and it represents the probability of survival at time  $t$ .

In most models of population dynamics the following time-dependent monomial is introduced

$$T(t) = \beta t^\alpha. \quad (15)$$

The expression (15) for  $T$  contains the real-valued parameters  $\alpha > 0$  and  $\beta > 0$ . We can think of  $T(t)$  as the time measured by a *nonlinear clock*<sup>107</sup>.  $T(t)$  is regularly used in the definition of the survival function  $S(t)$  which becomes an implicit function of time based on the dependence of  $S$  on  $T = T(t)$ , i.e.  $S = S(T)$ .

The survival function of the *Weibull model* is then defined according to

$$S = \exp(-T), \quad (16)$$

and it represents a stretched exponential function in time.

Popular models for empirical data that exhibit power-law tails include the Log-Logistic model<sup>108</sup>, with survival function given by

$$S = \frac{1}{1 + T}, \quad (17)$$

the Burr type XII or Singh-Maddala model<sup>109</sup>, with survival function given by

$$S = \frac{1}{(1 + T)^p}, \quad p > 0, \quad (18)$$

and the Dagum model<sup>110</sup>, with survival function given by

$$S = 1 - \frac{T^p}{(1 + T)^p}, \quad p > 0. \quad (19)$$

Various other postulates can be used for the analytical expression of the survival function. However, for physical applications it is extremely important to identify physical mechanisms or first principles which lead to such expressions for the survival function.

The  $\kappa$ -deformed statistical model can be viewed as a one-parameter generalization of the Weibull model, obtained by replacing the ordinary exponential  $\exp(t)$  in the definition (16) of the survival function by the  $\kappa$ -deformed exponential  $\exp_\kappa(t)$ . Then, using the Weibull dependence for  $T$  given by Eq. (15), we obtain the following expression for the  $\kappa$ -deformed survival function  $S_\kappa = S_\kappa(t)$

$$S_\kappa = \exp_\kappa(-T), \quad (20)$$

or more explicitly

$$S_\kappa(t) = \exp_\kappa(-\beta t^\alpha), \quad \alpha > 0, \beta > 0. \quad (21)$$

The  $\kappa$ -deformed survival function  $S_\kappa(t)$  reduces to the ordinary survival function of the Weibull model in the  $\kappa \rightarrow 0$  limit. The rate equation obeyed by  $S_\kappa$  is expressed in terms of  $T$  as a linear first-order ODE

$$\frac{dS_\kappa}{dT} = -\frac{1}{\sqrt{1 + \kappa^2 T^2}} S_\kappa, \quad (22)$$

with initial condition  $S_\kappa(0) = 1$ . The ODE (22) represents an interpolation between the rate equation of the exponential model and that of the Pareto model; moreover, the functional form in Eq. (22) is dictated by the first principles of special relativity.

The most important feature of the function  $S_\kappa(t)$  is that it continuously interpolates between a power-law tail for large  $t \gg 1$ , i.e.,

$$S_\kappa(t) \sim (2\kappa\beta)^{-1/\kappa} t^{-\alpha/\kappa}, \quad (23)$$

and exponential dependence

$$S_\kappa(t) \sim \exp(-\beta t^\alpha), \quad (24)$$

for  $t \ll 1$ .

The lifetime distribution function  $F_\kappa = F_\kappa(t)$  is given by the expression

$$F_\kappa(t) = 1 - \exp_\kappa(-\beta t^\alpha), \quad (25)$$

while the pdf  $f_\kappa = f_\kappa(t)$ , defined  $f_\kappa = dF_\kappa/dt$ , is given by

$$f_\kappa(t) = \frac{\alpha \beta t^{\alpha-1}}{\sqrt{1 + \kappa^2 \beta^2 t^{2\alpha}}} \exp_\kappa(-\beta t^\alpha). \quad (26)$$

The rate equation for the survival function assumes the form of the first-order linear ODE

$$\frac{dS_{\kappa}(t)}{dt} = -\lambda_{\kappa}(t) S_{\kappa}(t), \quad (27)$$

where  $S_{\kappa}(0) = 1$ , and  $\lambda_{\kappa} = \lambda_{\kappa}(t)$  is the *hazard function* (*hazard rate*) defined through

$$f_{\kappa}(t) = \lambda_{\kappa}(t) S_{\kappa}(t). \quad (28)$$

Hence, the hazard function assumes the expression

$$\lambda_{\kappa}(t) = \frac{\alpha \beta t^{\alpha-1}}{\sqrt{1 + \kappa^2 \beta^2 t^{2\alpha}}}. \quad (29)$$

The cumulative hazard function  $\Lambda_{\kappa} = \Lambda_{\kappa}(t)$  is defined by means of the integral

$$\Lambda_{\kappa}(t) = \int_0^t \lambda_{\kappa}(u) du, \quad (30)$$

and is linked with  $S_{\kappa}$  through

$$\Lambda_{\kappa}(t) = -\ln S_{\kappa}(t). \quad (31)$$

After taking into account that the  $\kappa$ -exponential function (10) can also be expressed in the form

$$\exp_{\kappa}(x) = \exp\left(\frac{1}{\kappa} \operatorname{arcsinh}(\kappa x)\right), \quad (32)$$

the cumulative hazard function  $\Lambda_{\kappa} = \Lambda_{\kappa}(t)$  assumes the following explicit expression

$$\Lambda_{\kappa}(t) = \frac{1}{\kappa} \operatorname{arcsinh}(\kappa \beta t^{\alpha}), \quad (33)$$

and in the  $\kappa \rightarrow 0$  limit reduces to the standard Weibull cumulative hazard function  $\Lambda(t) = \beta t^{\alpha}$ . It is important to note that in the standard Weibull model the cumulative hazard function coincides with the function  $T(t)$ .

Finally, the *quantile function*  $Q_{\kappa}$  is defined as the inverse of the survival function follows  $S_{\kappa} = S_{\kappa}(t)$ . By expressing the quantile function in the form  $t = Q_{\kappa}(S_{\kappa})$ , one easily obtains

$$Q_{\kappa}(u) = \left(-\frac{1}{\beta} \ln_{\kappa}(u)\right)^{1/\alpha}, \quad (34)$$

where the  $\kappa$ -logarithm  $\ln_{\kappa}(u)$  is the inverse function of  $\exp_{\kappa}(u)$ , i.e.  $\ln_{\kappa}(\exp_{\kappa}(u)) = \exp_{\kappa}(\ln_{\kappa}(u)) = u$  and is given by

$$\ln_{\kappa}(u) = \frac{u^{\kappa} - u^{-\kappa}}{2\kappa}. \quad (35)$$

After observing that the function  $\ln_{\kappa}(u)$  approaches the function  $\ln(u)$  in the  $\kappa \rightarrow 0$  limit, it follows that in the same limit  $Q_{\kappa}(u)$  reduces to the quantile function of the standard Weibull model.

Regarding the meaning of the parameters  $\alpha$ ,  $\beta$  and  $\kappa$ , we observe that the first two are the same as in the Weibull model and therefore have the same meaning. In particular,  $\alpha$  is the *shape parameter* while  $\beta$  is linked to the *scale parameter*  $\tau$  through the relationship  $\beta = \tau^{-1/\alpha}$ . The deformation parameter  $\kappa$ , on the other hand, is linked to the Pareto exponent  $n$  of the survival function through the simple relationship  $n = \alpha/\kappa$  and to the Pareto exponent  $p$  of the pdf via  $p = 1 + \alpha/\kappa$ ; the former follows from the asymptotic dependence of the survival function given by Eq. (23), while the latter from the fact that  $f(t) = -dS(t)/dt$ .

The mode of the probability density function, i.e. the time  $t_M$  where the pdf attains its maximum value can be easily obtained analytically as a function of the free parameters  $\alpha$ ,  $\beta$  and  $\kappa$ . More explicitly, after setting the first time derivative of the probability density function equal to zero we obtain the following biquadratic equation

$$T_M^4 \kappa^4 - [\alpha^2 T_M^4 + 2(\alpha - 1) T_M^2] \kappa^2 + (\alpha - 1)^2 - \alpha^2 T_M^2 = 0, \quad (36)$$

with  $T_M = \beta t_M^{\alpha}$ .

This equation can be solved to obtain  $T_M$  as a function of  $\alpha$  and  $\kappa$ . It can also be solved to determine  $\kappa$  as a function of  $\alpha$  and  $T_M$ , i.e.,

$$\kappa = \frac{1}{T_M} \sqrt{\alpha - 1 + \frac{1}{2} \alpha^2 T_M^2 - \alpha T_M \sqrt{\alpha + \frac{1}{4} \alpha^2 T_M^2}}. \quad (37)$$

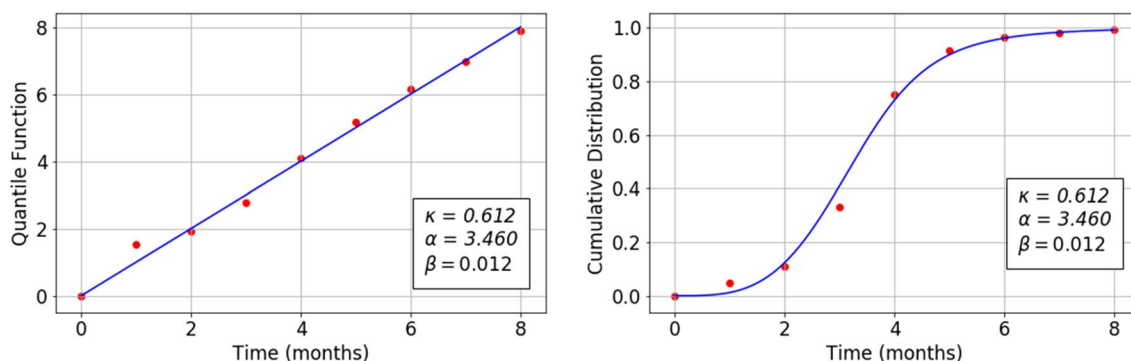
This last expression is very useful in the analysis of empirical data.

Finally, the  $\kappa$ -deformed statistical model presented above, which represents a one-parameter continuous deformation of the Weibull model, has been successfully applied in econophysics for the analysis of personal income distribution<sup>95</sup> and in seismology<sup>104</sup>. We believe that the applicability of the model to such complex dynamical systems which are quite different from each other suggests that the model involves universal features.



Month	Dead people	Population of Florence
May	600	59400
June	700	58700
July	2700	56000
August	5000	51000
September	2000	49000
October	600	48400
November	200	48200
December	100	48100

**Table 1.** Temporal distribution of plague victims and population in Florence during the year 1417<sup>24</sup>.



**Figure 1.** Theoretical (continuous curve) and empirical (dots) plots of the quantile function (left) and the cumulative distribution function (right) versus time for the 1417 Florence plague epidemic. The theoretical curves are based on Eqs. (34) and (25), respectively.

## Results

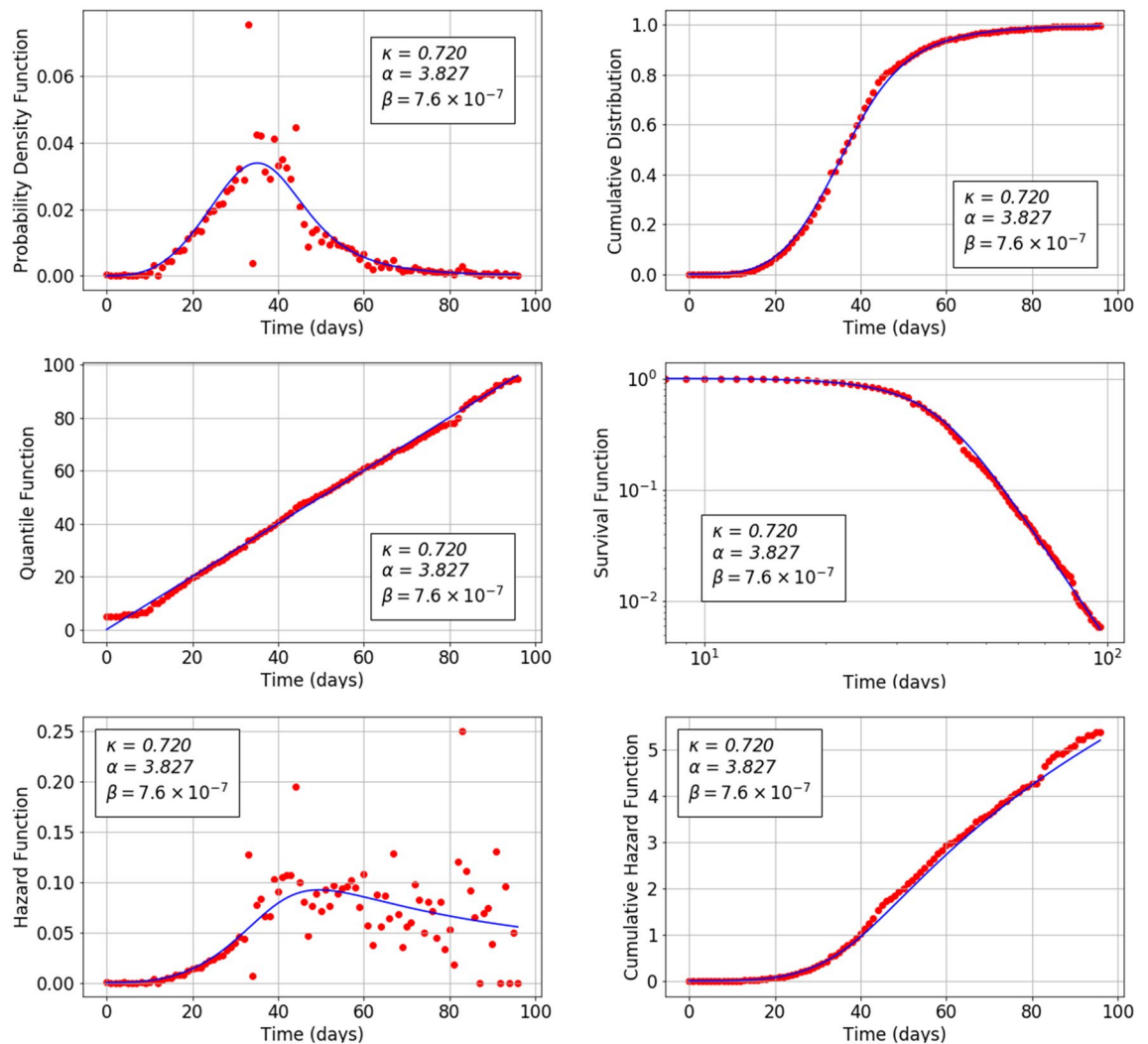
The purpose of this section is the validation of the  $\kappa$ -Weibull statistical model which is described in detail in the “Methods” section. The validation is performed by applying the model to various pandemic data. An intriguing first application regards the analysis of the number of deaths during the plague pandemic that ravaged the city of Florence in the XV century.

The Bubonic Plague or Black Death arrived in Europe in 1348 and in Italy in the spring of the same year<sup>22,23</sup>. It should be noted that the plague inspired Boccaccio to write his famous nouvelle *Decameron* only a couple of years after the end of the pandemic. Its consequence was the death of around 25–50% of Europe’s population by 1351. The pandemic is believed to have started in China and came to Europe via the trading routes through Asia and the Black Sea. In the year 1417, Florence had a population of 60,000 inhabitants and the Italian city’s Grain Office maintained a series of Books of Dead which recorded the number of human deaths caused by the bubonic plague. Florentines contracted the disease through contact with infected black rats and the fleas that they carried<sup>23</sup>. From a thermodynamic point of view, deaths which are due to the spread of an epidemic can be considered as an irreversible process inside an open system. Herein we show how the evolution of the plague in relation to the time and the number of deceased people can be explained using the  $\kappa$ -deformed statistics. In Table 1 the statistical data of the 1417 plague in Florence are summarised as a function of time<sup>24</sup>.

We consider the *death occurrence probability* as the ratio between the monthly number of deaths (given by the second column of the Table 1) and the cumulative number of deaths. We assume that the latter is equal to the total number of deaths that were recorded between May and December increased by an additional 100 deaths; we hypothesize that the latter occurred in the months following December, raising the total number of deaths to  $N = 12000$ .

Figure 1 shows the quantile function  $Q_\kappa$  (left) and the cumulative distribution function  $F_\kappa$  (right), versus the time  $t$  for the Florence epidemic. The theoretical quantile function of the  $\kappa$ -deformed model, defined in Eq. (34), is just the bisectrix of the plane represented by the continuous straight line. The continuous sigmoid curve represents the theoretical cumulative distribution function as given by Eq. (25). The empirical data related to the two functions (marked on the plots by dots) are deduced from the data in Table 1. The optimal fit parameters have the following values:  $\kappa = 0.612$ ,  $\alpha = 3.460$  and  $\beta = 0.012$ . It is clear that the model describes the empirical data very well. Deviations of the empirical quantiles from the theoretical curves are likely due to detection errors since as it is evidenced in Table 1, the number of monthly deaths is approximated with an error margin of the order of 100 units. This statistical error is responsible for the dispersion of empirical quantiles around the bisectrix. Hence, the optimal parameter estimate  $\kappa = 0.612$  clearly shows the difference (supported by the data) between the optimal  $\kappa$ -deformed model and the standard Weibull model corresponding to value  $\kappa = 0$ .

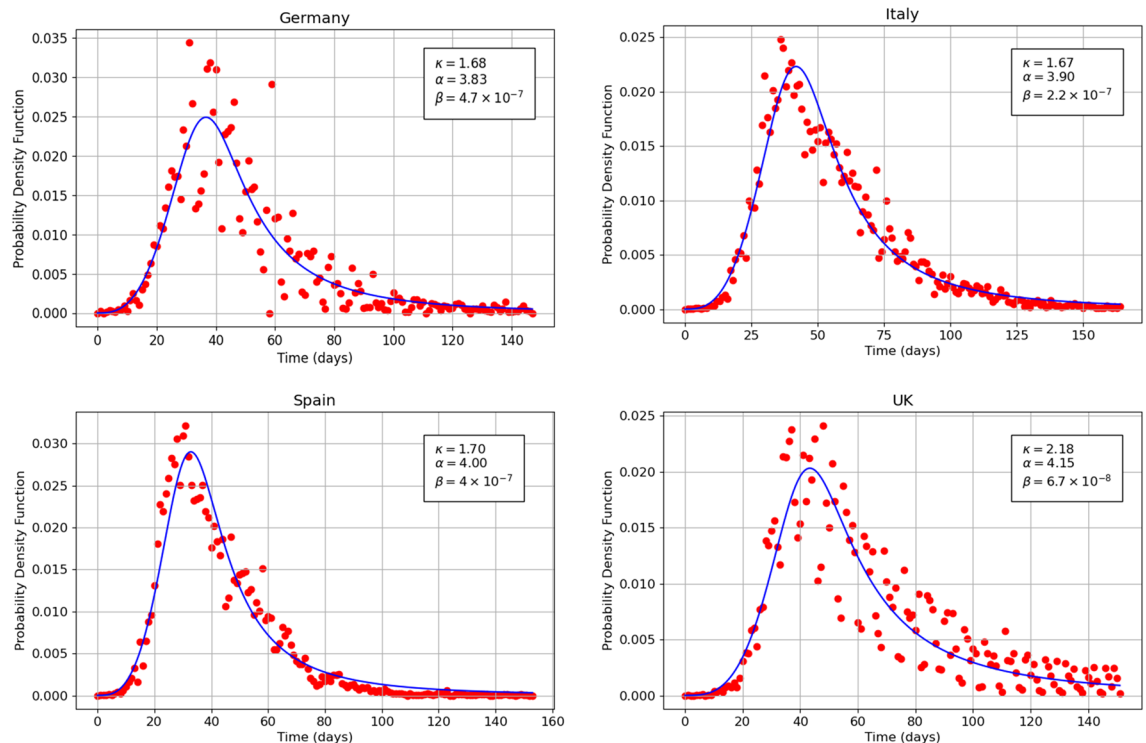




**Figure 2.** Theoretical (continuous curve) and empirical (dots) plots of the probability density function (top left), cumulative distribution function (top right), quantile function (middle left), survival function (middle right), hazard function (bottom left) and cumulative hazard function (bottom right) versus time for the Covid-19 mortality data related to China 2020 epidemic. The theoretical curves are based on Eqs. (26), (25), (34), (21), (29) and (33) respectively.

As second application we analyze the data of Covid-19 mortality in China during the winter and spring of 2020<sup>25,26</sup>. The Chinese data on SARS-CoV-2 are important, as they represent the first statistical record for the entire evolutionary cycle of the COVID-19 pandemic in the world. It is important to note that the first and only cycle of Covid-19 ended in China with a total of 3,342 deaths around April 16, 2020. After April 17, 2020 while the Covid-19 cycle in China had already ended, the authorities reported 1290 additional deaths that occurred outside the hospitals during the entire cycle without disclosing their temporal distribution. For this reason the present analysis is applied to the 3,342 deaths which were reported by April 16, 2020.

Figure 2 reports the probability density function  $f_\kappa$  Eq. (26), the cumulative distribution function  $F_\kappa$  Eq. (25), the quantile function  $Q_\kappa$  Eq. (34), the survival function  $S_\kappa$  Eq. (21), the hazard function  $\lambda_\kappa$  Eq. (29) and the cumulative hazard function  $\Lambda_\kappa$  Eq. (33) versus the time  $t$  (measured in days) for the COVID-19 pandemic in China. The best fit parameters have the values:  $\kappa = 0.720$ ,  $\alpha = 3.827$  and  $\beta = 7.6 \cdot 10^{-7}$ . The plots in this figure confirm the goodness of fit of the  $\kappa$ -deformed model. In particular, there is no systematic deviation of the empirical quantiles from the theoretical linear trend while the statistical nature of data leads to slight dispersion around the bisectrix. Furthermore the survival function  $S_\kappa$  versus time  $t$  is shown on a log-log plot in order to better explore differences between the empirical data and the theoretical predictions, especially in the tail of the distribution. The almost linear decay of the tail in this log-log plot is remarkable, showing clear signs that the Chinese Covid-19 data follows the Pareto power law in the distribution tail. The dispersion around the theoretical curves of the empirical data for both  $f_\kappa$  and  $\lambda_\kappa$  reflects statistical fluctuations that are independent on the adopted theoretical model. These fluctuations are averaged out and become less evident in the cumulative functions  $F_\kappa$ ,  $S_\kappa$ ,  $\Lambda_\kappa$  and  $Q_\kappa$ .



**Figure 3.** Theoretical (continuous curve) and empirical (dots) plots of the probability density function versus time for the Covid-19 mortality data related to Germany (top left), Italy (top right), Spain (bottom left) and United Kingdom (bottom right) 2020 pandemic. The theoretical curves are based on Eq. (26).

In the following we focus on the pandemic in Europe; in particular, we analyze the mortality Covid-19 data of Germany, Italy, Spain and United Kingdom, during the spring of 2020<sup>25,26</sup>. In these European countries the virus began to spread later than in China while the mortality exceeded the Chinese rate by more than 10 times.

Figure 3 shows the probability density function  $f_k(t)$  versus time related to Covid-19 mortality daily data for Germany, Italy, Spain and United Kingdom. The theoretical curve is given by Eq. (26). The parameter  $\alpha$  for the four countries assumes the values 3.83, 3.90, 4.00 and 4.15 respectively. The parameter  $\kappa$  for the four countries assumes the values 1.68, 1.67, 1.70 and 2.185 respectively. Finally the parameter  $\beta$  for the four countries assumes the values  $4.7 \cdot 10^{-7}$ ,  $2.2 \cdot 10^{-7}$ ,  $4.00 \cdot 10^{-7}$  and  $6.7 \cdot 10^{-8}$  respectively. The comparison of the above optimal parameters for the four countries is performed below.

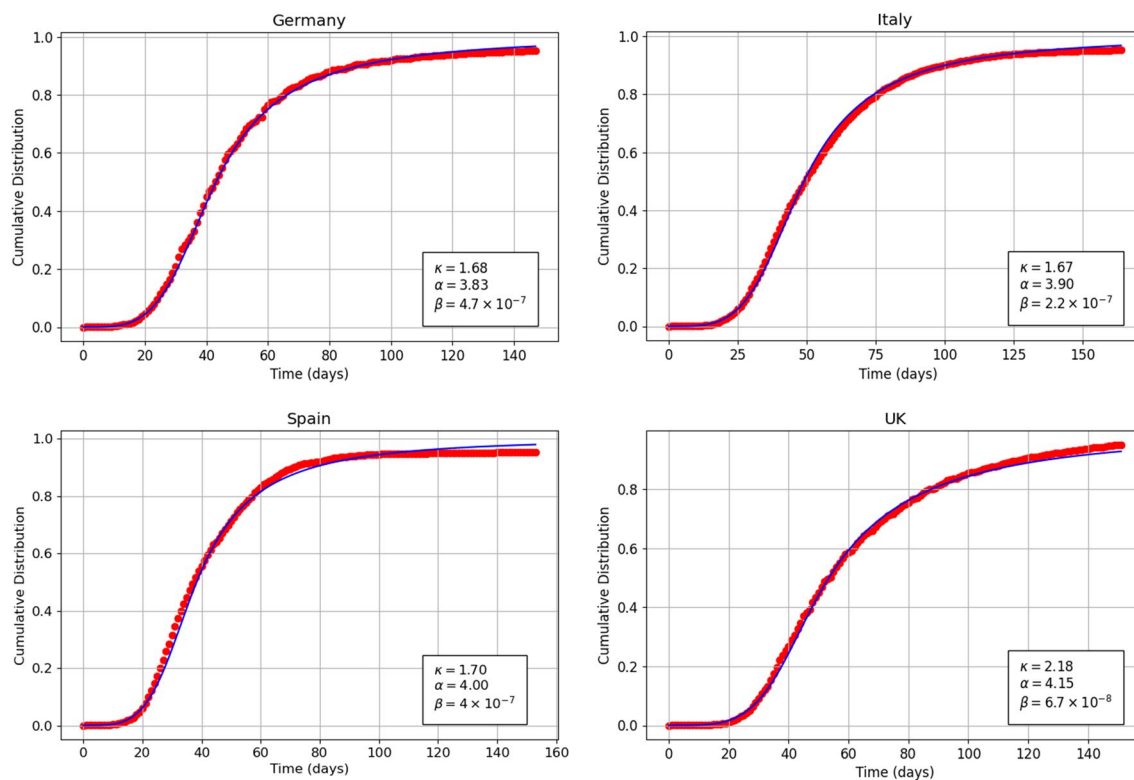
Figure 4 shows the cumulative distribution function  $F_k(t)$  versus time related to Covid-19 mortality data for Germany, Italy, Spain and United Kingdom. The theoretical curve is given by Eq. (25). The optimal values of the parameters  $\alpha$ ,  $\kappa$  and  $\beta$  for the four countries are the same as in Fig. 3. It is worth noting that the statistical fluctuations of the daily data which are evident in Fig. 3 are absorbed in the representation of the cumulative data shown in Fig. 4.

Figure 5 displays the survival function  $S_k(t)$  versus time on a log-log plot related to Covid-19 mortality data for Germany, Italy, Spain and United Kingdom. The continuous curve is the theoretical function defined in Eq. (21). The optimal values of the parameter  $\alpha$ ,  $\kappa$  and  $\beta$  for the four countries are the same as in Fig. 3. In this log-log plot, a remarkable agreement between the theoretical predictions and the empirical data clearly emerges. Furthermore, the almost linear decay of the tail of the survival function shown in this plot indicates a Pareto power-law tail for all the countries studied here, in agreement with the  $\kappa$ -deformed model.

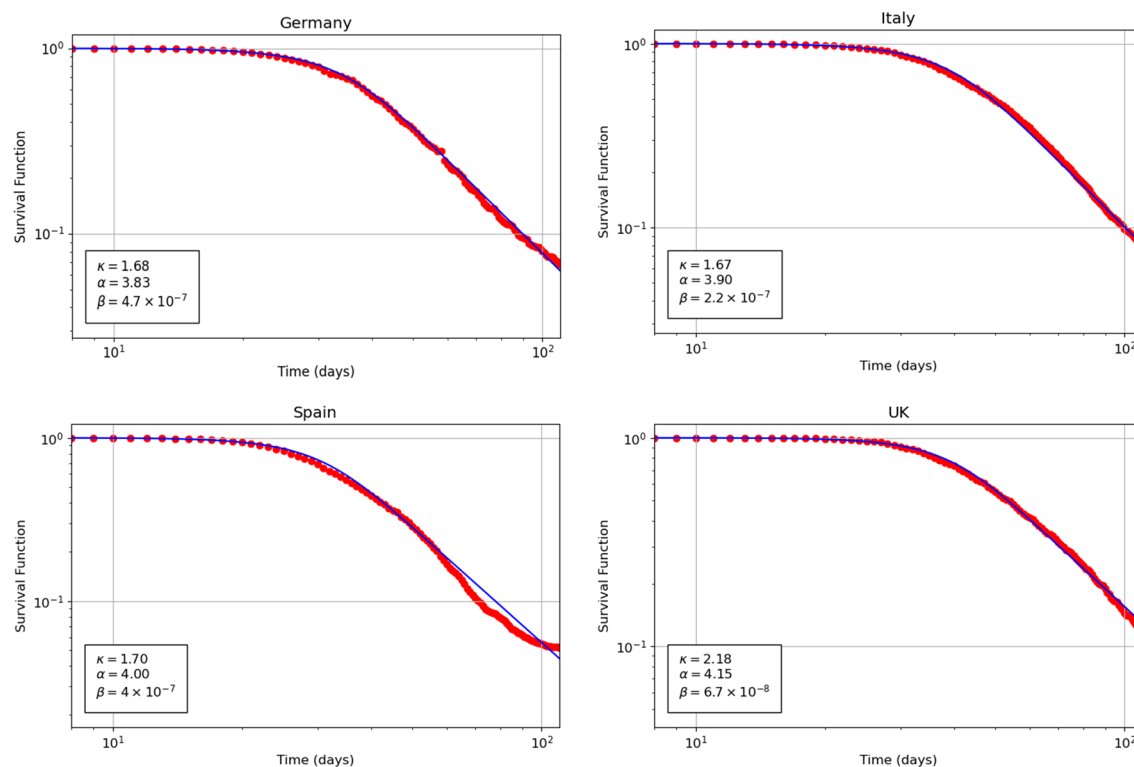
For the purpose of comparing the Covid-19 mortality data from China, Germany, Italy, Spain and United Kingdom we focus on the probability density function. First, we discuss the growth in mortality (i.e., the left tail) which is due to the initial phase of the virus spread. This is followed by a discussion of the final phase of the studied cycle, during which the decrease in the virus spread is described by the right tail of the curve.

The asymptotic behaviour of  $f_k(t)$  for  $t \rightarrow 0$  (left tail) is given by  $f_k(t) \propto t^\alpha$  where  $\tau = \beta^{-1/\alpha}$  is the *Weibull scale parameter*. The shape parameter for China, Germany, Italy, Spain and United Kingdom has the values 3.83, 3.83, 3.90, 4.00 and 4.15 respectively. These estimates of  $\alpha$  are very close in value, with an average value equal to  $3.94 \pm 0.2$ . It can therefore be concluded that for the five countries analyzed, the initial growing phase of the Covid-19 mortality presents quite similar dynamic behavior.

The asymptotic behaviour of  $f_k(t)$  for  $t \rightarrow \infty$  (right tail) is given by  $f_k(t) \propto t^{-p}$  where  $p = 1 + \alpha/\kappa$  is the pdf's Pareto parameter (tail exponent). This parameter which governs the behavior of the right tail of the distribution takes the values 6.31, 3.28, 3.33, 3.35 and 2.83 for China, Germany, Italy, Spain and United Kingdom respectively. It can be observed that the dynamics of the final phase of the first cycle of the epidemic is essentially the same for the three continental European countries in the study. For these countries the Pareto parameter is approximately equal to  $3.32 \pm 0.04$ . For the United Kingdom, on the other hand, the Pareto parameter is lower



**Figure 4.** Theoretical (continuous curve) and empirical (dots) plots of the cumulative distribution function versus time for the Covid-19 mortality data related to Germany (top left), Italy (top right), Spain (bottom left) and United Kingdom (bottom right) 2020 pandemic. The theoretical curves are based on Eq. (25).



**Figure 5.** Theoretical (continuous curve) and empirical (dots) plots of the survival function versus time for the Covid-19 mortality data related to Germany (top left), Italy (top right), Spain (bottom left) and United Kingdom (bottom right) 2020 pandemic. The theoretical curves are based on Eq. (21).

(equal to 2.83) indicating that the tail is thicker, i.e., that mortality is more persistent over time. On the contrary, the Pareto index is higher (equal to 6.31) for the Chinese data which indicates a faster decay of the mortality curve than the respective European curves. The differences in the Pareto parameters between the different countries are likely to reflect differences in the response strategies, in the compliance of the citizens to restrictive measures that can curb the spread of the virus, as well as in the readiness and quality of the health care system including emergency treatment units.

## Discussion

Both the plague data from the pandemic of 1417 in Florence as well as the Covid-19 data of the 2020 pandemic from China, Germany, Italy, Spain and United Kingdom have been analyzed by means of the proposed  $\kappa$ -Weibull model, to obtain information about the spreading dynamics of these deadly disease outbreaks.

It is worth recalling that the Florence plague data are approximate and relatively sparse, and thus they cannot be used to reliably test a statistical model. However, the Florence data are of great historical importance as they represent the first quantitative and regularly collected record of a pandemic. The first two figures in this paper undoubtedly show that at least the cumulative Florence data are compatible with the proposed  $\kappa$ -deformed statistical model.

On the other hand, the Chinese data set on SARS-CoV-2 are important, as they represent the first statistical record for the entire evolutionary cycle of the COVID-19 pandemic in the world. The Chinese mortality data have been successfully used to perform a first validation test of the  $\kappa$ -deformed model.

It is important to note that the first and only (so far) cycle of Covid-19 ended in China with a total of 3,342 deaths. This is a relatively low mortality compared to other countries where the number of deaths was decidedly higher and mortality figures rose to a few dozen times that of China. A second noteworthy fact is that on April 17, 2020 while the Covid-19 cycle in China had already ended, the Chinese authorities reported 1290 additional deaths that occurred outside hospitals during the entire cycle without providing any information regarding their temporal distribution.

It is therefore evident that the data pertaining to the 3,342 deaths caused by Covid-19 in China are partial. For this reason, the validation of the  $\kappa$ -deformed model with complete data from other countries affected by Covid-19 is extremely important. Thus, we focused on Europe and analyzed the Covid-19 data of Italy and Spain, where the virus began to spread towards the beginning of March 2020, as well as the data from Germany and the United Kingdom, where the virus began to spread two weeks later. These data from the European countries and China differ with respect to the total number of deaths, which in the case of Germany is about 3 times the number of Chinese deaths while for Italy, Spain and United Kingdom the mortality is about 10 times that of China. Very good agreement has been obtained by fitting the  $\kappa$ -Weibull model to the data from these European countries, thereby providing further and more reliable validation of the theoretical model.

The  $\kappa$ -deformed, three-parameter model admits simple analytical forms for all the main statistical functions (probability density function, cumulative distribution function, survival function, quantile function, hazard function, cumulative hazard function) and can therefore be easily applied to Covid-19 data. Furthermore this model preserves all the universality properties already present in the original Weibull model. The results obtained by applying the  $\kappa$ -deformed model to mortality data from Covid-19 pertaining to the first cycle of the pandemic in China, Germany, Italy, Spain and United Kingdom, suggest that the  $\kappa$ -deformed model could also be used to model infection data from the first cycle of the Covid-19 pandemic as well as for the analysis of the second pandemic cycle which has already started in August 2020.

Received: 11 June 2020; Accepted: 23 October 2020

Published online: 17 November 2020

## References

1. Martin, B. R. *Statistics for Physical Science* (Academic Press, Boston, 2012).
2. Sornette, D. *Critical Phenomena in Natural Sciences: Chaos, Fractals, Self-organization and Disorder: Concepts and Tools* (Springer, Berlin, 2006).
3. West, G. B. *Scale: the Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies* (Penguin, New York, 2017).
4. Christakos, G. & Hristopulos, D. *Spatiotemporal Environmental Health Modelling* (Springer Science and Business Media, New York, 1998).
5. Daley, D. J. & Gani, J. *Epidemic Modelling: An Introduction, Cambridge Studies in Mathematical Biology* (Cambridge University Press, New York, 1999).
6. Madhav, N. *et al.* Chapter 17: Pandemics: Risks, impacts, and mitigation. In *Disease Control Priorities: Improving Health and Reducing Poverty* 3rd edn (eds Jamison, D. T. *et al.*) (The World Bank, Washington, 2017).
7. Porta, M. *A Dictionary of Epidemiology* 6th edn. (Oxford University Press, Oxford, 2014).
8. Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
9. Morse, S. S. Factors in the emergence of infectious diseases. *Emerg. Infect. Dis.* **1**, 7–15 (1995).
10. Woolhouse, M. E. J. & Gowtage-Sequeria, S. Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* **11**, 1842–1847 (2005).
11. Sands, P., Mundaca-Shah, C. & Dzau, V. J. The neglected dimension of global security—a framework for countering infectious-disease crises. *N. Engl. J. Med.* **374**, 1281–1287 (2016).
12. Lucia, U. Statistical approach of the irreversible entropy variation. *Phys. A* **387**, 3454–3460 (2008).
13. Lucia, U. Irreversibility, entropy and incomplete information. *Phys. A* **388**, 4025–4033 (2009).
14. Lucia, U., Deisboeck, T. & Grisolia, G. Entropy-based pandemic forecasting. *Front. Phys.* **8**, 274 (2020).
15. Tsallis, C. & Tirnakli, U. Predicting COVID-19 peaks around the world. *Front. Phys.* **8**, 217 (2020).
16. Vasconcelos, G. L. *et al.* Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies. *MedRxiv* <https://doi.org/10.1101/2020.04.02.20051557> (2020).

17. Zlatić, V., Barjašić, I., Kadović, A., Štefančić, H. & Gabrielli, A. Bi-stability of SUDR+K model of epidemics and test kits applied to COVID-19. *arXiv [Preprint]* 2003.08479 (2020).
18. Piccolomini, E. L. & Zama, F. Preliminary analysis of COVID-19 spread in Italy with an adaptive SEIRD model. *arXiv [Preprint]* 2003.09909 (2020).
19. Bastos, S. & Cajueiro, D. Modeling and forecasting the early evolution of the Covid-19 pandemic in Brazil. *arXiv [Preprint]* 2003.14288v2 (2020).
20. Perc, M., Miksic, N. G., Slavinec, M. & Stožer, A. Forecasting COVID-19. *Front. Phys.* **8**, 127 (2020).
21. Pluchino, A. *et al.* A novel methodology for epidemic risk assessment: the case of COVID-19 outbreak in Italy. *arXiv [Preprint]* 2004.02739 (2020).
22. George, C. K. (Ed.). *Encyclopedia of Plague and Pestilence: From Ancient Times to the Present*. 3rd edn. (Facts on File, New York, 2008).
23. Cohn, S. K. J. The black death: the end of a paradigm. In *Power, Violence and Mass Death in Pre-Modern and Modern Times* (eds Canning, J. *et al.*) (Ashgate Publishing Limited, Aldershot, 2004).
24. Le Panta, L. D. *Epidemie Nella Storia Demografica Italiana [The Plagues in the Italian Demographic History]* (Loesher, Turin, 1986).
25. <https://www.worldometers.info/coronavirus/>.
26. <https://www1.nyc.gov/site/doh/covid/covid-19-data-archive.page>.
27. Chen, W.-Y. & Bokka, S. Stochastic modeling of nonlinear epidemiology. *J. Theor. Biol.* **234**, 455–470 (2005).
28. Zipf, G. K. *Selected Studies of the Principle of Relative Frequency in Language* (Harvard University Press, Boston, 1932).
29. Zipf, G. K. *The Psycho-biology of Language: An Introduction to Dynamic Philology* (Houghton-Mifflin Co., Boston, 1935).
30. Zipf, G. K. *National Unity and Disunity* (The Principia Press, Bloomington, 1940).
31. Zipf, G. K. *Human Behavior and the Principle of Least Effort* (Addison-Wesley, New York, 1949).
32. Wyllys, R. E. Empirical and theoretical bases of Zipf's law. *Library Trends* **30**, 53–64 (1981).
33. Perline, R. Zipf's law, the central limit theorem and the random division of the unit interval. *Phys. Rev. E* **54**, 220 (1996).
34. Okuyama, K., Takayasu, M. & Takayasu, H. Zipf's law in income distribution of companies. *Phys. A* **269**, 125–131 (1999).
35. Li, W. Zipf's law everywhere. *Glottometrics* **5**, 14–21 (2002).
36. Newman, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351 (2005).
37. Benguigui, L. & Blumenfeld-Lieberthal, E. The end of a paradigm: is Zipf's law universal? *J. Geogr. Syst.* **13**, 87–100 (2011).
38. Piantadosi, S. T. Zipf's word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.* **21**, 1112–1130 (2014).
39. Kleiber, C. & Kotz, S. *Distributions in Economics and Actuarial Sciences* (John Wiley & Sons, Hoboken, 2003).
40. Perline, R. Strong, weak and false inverse power laws. *Stat. Sci.* **20**, 68–88 (2005).
41. Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703 (2009).
42. Pareto, V. *Ecrits sur la courbe de la repartition de la richesse* (1896). In *Ouvres complètes de Vilfredo Pareto publiées sous la direction de Giovanni Busino* (Librairie Droz, Genève, 1965).
43. Dagum, C. Wealth distribution models: Analysis and applications. *Stat. LXVI* **2**, 235–268 (2006).
44. Pareto, V. *Cours d'Economie Politique* (Rouge, Lausanne, 1897).
45. Toda, A. A. The double power law in income distribution: Explanations and evidence. *J. Econ. Behav. Organ.* **84**, 364–381 (2012).
46. Kaniadakis, G. Statistical mechanics in the context of special relativity. *Phys. Rev. E* **66**, 17 (2002).
47. Kaniadakis, G. Maximum entropy principle and power-law tailed distributions. *Eur. Phys. J. B* **70**, 3–13 (2009).
48. Kaniadakis, G., Scarfone, A. M., Sparavigna, A. & Wada, T. Composition law of kappa-entropy for statistically independent systems. *Phys. Rev. E* **95**, 052112 (2017).
49. Silva, R. The relativistic statistical theory and Kaniadakis entropy: an approach through a molecular chaos hypothesis. *Eur. Phys. J. B* **54**, 499 (2006).
50. Naudts, J. Deformed exponentials and logarithms in generalized thermostats. *Phys. A* **316**, 323 (2002).
51. Topsoe, F. Entropy and equilibrium via games of complexity. *Phys. A* **340**, 11 (2004).
52. Tempesta, P. Group entropies, correlation laws, and zeta functions. *Phys. Rev. E* **84**, 021121 (2011).
53. Scarfone, A. M. Entropic forms and related algebras. *Entropy* **15**, 624 (2013).
54. Souza, N. T. C. M., Anselmo, D. H. A. L., Silva, R., Vasconcelos, M. S. & Mello, V. D. Analysis of fractal groups of the type d-(m, r)-Cantor within the framework of Kaniadakis statistics. *Phys. Lett. A* **378**, 1691 (2014).
55. Scarfone, A. M. On the  $\kappa$ -deformed cyclic functions and the generalized Fourier series in the framework of the kappa-algebra. *Entropy* **17**, 2812 (2015).
56. Scarfone, A. M.  $\kappa$ -deformed Fourier transform. *Phys. A* **480**, 63 (2017).
57. Gomez, I. S. & Borges, E. Universality classes for the Fisher metric derived from relative group entropy. *Phys. A* **547**, 123827 (2020).
58. da Silva, J. L. E. & Ramos, R. The Lambert-Kaniadakis  $w_\kappa$  function. *Phys. Lett. A* **384**, 126175 (2020).
59. Wada, T. Thermodynamic stabilities of the generalized Boltzmann entropies. *Phys. A* **340**, 126 (2004).
60. Bento, E. P., Viswanathan, G. M., da Luz, M. G. E. & Silva, R. Third law of thermodynamics as a key test of generalized entropies. *Phys. Rev. E* **91**, 022105 (2015).
61. Santos, A. P., Silva, R., Alcaniz, J. S. & Anselmo, D. H. A. L. Kaniadakis statistics and the quantum H-theorem. *Phys. Lett. A* **375**, 352 (2011).
62. Ourabah, K. & Tribeche, M. Planck radiation law and Einstein coefficients reexamined in Kaniadakis kappa statistics. *Phys. Rev. E* **89**, 062130 (2014).
63. Lourek, I. & Tribeche, M. Thermodynamic properties of the blackbody radiation: A Kaniadakis approach. *Phys. Lett. A* **381**, 452 (2017).
64. Soares, B. B., Barboza, E. M. & Neto, J. Non-gaussian thermostats considerations upon the Saha equation. *Phys. A* **532**, 121590 (2019).
65. Ourabah, K., Hamici-Bendimerad, A. H. & Tribeche, M. Quantum Kaniadakis entropy under projective measurement. *Phys. Rev. E* **92**, 032114 (2015).
66. Ourabah, K., Hamici-Bendimerad, A. H. & Tribeche, M. Quantum entanglement and Kaniadakis entropy. *Phys. Scr.* **90**, 045101 (2015).
67. Lourek, I. & Tribeche, M. On the role of the kappa-deformed Kaniadakis distribution in nonlinear plasma waves. *Phys. A* **441**, 215 (2016).
68. Gougam, L. A. & Tribeche, M. Electron-acoustic waves in a plasma with a kappa-deformed Kaniadakis electron distribution. *Phys. Plasmas* **23**, 014501 (2016).
69. Chen, H., Zhang, S. X. & Liu, S. Q. The longitudinal plasmas modes of  $\kappa$ -deformed kaniadakis distributed plasmas. *Phys. Plasmas* **24**, 022125 (2017).
70. Lopez, R. A., Navarro, R. E., Pons, S. I. & Arandeda, J. A. Landau damping in Kaniadakis and Tsallis distributed electron plasmas. *Phys. Plasmas* **24**, 102119 (2017).
71. Saha, A. & Tamang, J. Qualitative analysis of the positron-acoustic waves in electron-positron-ion plasmas with kappa deformed Kaniadakis distributed electrons and hot positrons. *Phys. Plasmas* **24**, 082101 (2017).



72. Lourek, I. & Tribeche, M. Dust charging current in non equilibrium dusty plasma in the context of kaniadakis generalization. *Phys. A* **517**, 522–529 (2019).
73. Khalid, M. & Rahman, A.-U. Oblique ion acoustic excitations in a magnetoplasma having  $\kappa$ -deformed Kaniadakis distributed electrons. *Astr. Space Sc.* **365**, 75 (2020).
74. Guedes, G., Goncalves, A. C. & Palma, D. A. P. Doppler broadening function using the Kaniadakis distribution. *Ann. Nucl. Energy* **126**, 262 (2019).
75. de Abreu, W. V., Goncalves, A. C. & Martinez, A. S. Analytical solution for the Doppler broadening function using the Kaniadakis distribution. *Ann. Nucl. Energy* **126**, 262 (2019).
76. de Abreu, W. V. & Martinez, A. New analytical formulations for the Doppler broadening function and interference term based on Kaniadakis distributions. *Ann. Nucl. En.* **135**, 106960 (2020).
77. Shen, K.-M. Analysis on hadron spectra in heavy-ion collisions with a new non-extensive approach. *J. Phys. G* **46**, 105101 (2019).
78. Carvalho, J. C., Silva, R., Nascimento, J. D. & Medeiros, J. R. D. Power law statistics and stellar rotational velocities in the Pleiades. *Europhys. Lett.* **84**, 59001 (2008).
79. Carvalho, J. C., Nascimento, J. D., Silva, R. & Medeiros, J. R. D. Non-gaussian statistics and stellar rotational velocities of main sequence field stars. *Astrophys. J. Lett.* **696**, L48 (2009).
80. Carvalho, J., Silva, R., Nascimento, J., Soares, B. B. & Medeiros, J. R. D. Observational measurement of open stellar clusters: A test of Kaniadakis and Tsallis statistics. *Europhys. Lett.* **91**, 69002 (2010).
81. Cure, M., Rial, D. F., Christen, A. & Cassetti, J. A method to deconvolve stellar rotational velocities. *Astron. Astrophys.* **564**, A85 (2014).
82. Abreu, E. M. C., Neto, J. A., Barboza, E. M. & Nunes, R. C. Jeans instability criterion from the viewpoint of Kaniadakis statistics. *EPL* **114**, 55001 (2016).
83. Abreu, E. M. C., Neto, J. A., Barboza, E. M. & Nunes, R. C. Tsallis and Kaniadakis statistics from the viewpoint of entropic gravity formalism. *Int. J. Mod. Phys.* **32**, 1750028 (2017).
84. Chen, H., Zhang, S. X. & Liu, S. Q. Jeans gravitational instability with kappa-deformed Kaniadakis distribution. *Chin. Phys. Lett.* **34**, 075101 (2017).
85. Abreu, E. M. C., Neto, J. A., Mendes, A. C. R. & Bonilla, A. Tsallis and Kaniadakis statistics from a point of view of the holographic equipartition law. *EPL* **121**, 45002 (2018).
86. Abreu, E. M. C., Neto, J. A., Mendes, A. C. R., Bonilla, A. & de Paula, R. M. Cosmological considerations in Kaniadakis statistics. *EPL* **124**, 30003 (2018).
87. Abreu, E. M. C., Neto, J. A., Mendes, A. C. R. & de Paula, R. M. Loop quantum gravity Immirzi parameter and the Kaniadakis statistics. *Chaos, Solitons Fractals* **118**, 307–310 (2019).
88. Oreste, P. & Spagnoli, G. Statistical analysis of some main geomechanical formulations evaluated with the Kaniadakis exponential law. *Geomech. Geoeng.* **13**, 139 (2018).
89. Oreste, P. & Spagnoli, G. Relation water content ratio-to-liquidity index versus the Atterberg limits ratio evaluated with the Kaniadakis exponential law. *Geomech. Geoeng.* **14**, 148 (2019).
90. Souza, N. T. C. M., Anselmo, D. H. A. L., Silva, R., Vasconcelos, M. S. & Mello, V. D. A kappa-statistical analysis of the Y-chromosome. *EPL* **108**, 28004 (2014).
91. Costa, M. O., Silva, R., Anselmo, D. H. A. L. & Silva, J. R. P. Analysis of human DNA through power-law statistics. *Phys. Rev. E* **99**, 022112 (2019).
92. Macedo-Filho, A., Moreira, D. A., Silva, R. & da Silva, L. R. Maximum entropy principle for Kaniadakis statistics and networks. *Phys. Lett. A* **377**, 842 (2013).
93. Stella, M. & Brede, M. A kappa-deformed model of growing complex networks with fitness. *Phys. A* **407**, 360–368 (2014).
94. Lei, B. & Fan, J.-L. Adaptive Kaniadakis entropy thresholding segmentation algorithm based on particle swarm optimization. *Soft Comp.* **24**, 7305–7318 (2020).
95. Clementi, F., Gallegati, M. & Kaniadakis, G. A model of personal income distribution with application to Italian data. *Empirical Econ.* **39**, 559–591 (2011).
96. Bertotti, M. & Modenese, G. Exploiting the flexibility of a family of models for taxation and redistribution. *Eur. Phys. J. B* **85**, 261 (2012).
97. Modenese, G. Common origin of power-law tails in income distributions and relativistic gases. *Phys. Lett. A* **380**, 29–31 (2016).
98. Bertotti, M. L. & Modanesi, G. Statistics of binary exchange of energy or money. *Entropy* **19**, 465 (2017).
99. Vallejos, A., Ormazabal, I., Borotto, F. A. & Astudillo, H. F. A new kappa-deformed parametric model for the size distribution of wealth. *Phys. A* **14**, 819–829 (2019).
100. Trivellato, B. The minimal  $\kappa$ -entropy martingale measure. *Int. J. Theor. Appl. Finance* **15**, 1250038 (2012).
101. Trivellato, B. Deformed exponentials and applications to finance. *Entropy* **15**, 3471 (2013).
102. Tapiero, O. J. A maximum (non-extensive) entropy approach to equity options bid-ask spread. *Phys. A* **392**, 3051 (2013).
103. Moretto, E., Pasquali, S. & Trivellato, B. A non-Gaussian option pricing model based on Kaniadakis exponential deformation. *Eur. Phys. J. B* **90**, 179 (2017).
104. Hristopulos, D. T., Petrakis, M. P. & Kaniadakis, G. Finite-size effects on return interval distributions for weakest-link-scaling systems. *Phys. Rev. E* **89**, 052142 (2014).
105. Hristopulos, D. T., Petrakis, M. P. & Kaniadakis, G. Weakest-link scaling and extreme events in finite-sized systems. *Entropy* **17**, 1103–1122 (2015).
106. da Silva, S. L. E. F., Carvalho, P. T. C. & Corso, G. Full-waveform inversion based on Kaniadakis statistics. *Phys. Rev. E* **101**, 053311 (2020).
107. Omalley, D. & Cushman, J. H. Fractional Brownian motion run with a nonlinear clock. *Phys. Rev. E* **82**, 032102 (2010).
108. Collett, D. *Modelling Survival Data in Medical Research* (CRC Press, New York, 2003).
109. Singh, S. & Maddala, G. A function for the size distribution of incomes. *Econometrica* **44**, 963–970 (1976).
110. Dagum, C. A new model of personal income distribution: Specification and estimation. *Economie Appliquée* **30**, 413–437 (1977).

## Author contributions

G.K. developed the statistical model and U.L. conceived the idea to adopt the model in the analysis of pandemic data. All authors provided critical feedback and helped shape the research, analysis and manuscript.

## Competing interests:

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to G.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020