The K indicator epidemic model follows the Gompertz curve

Yutaka Akiyama
Department of Computer Science, School of Computing
Tokyo Institute of Technology, Japan
E-mail: akiyama@c.titech.ac.jp

(revision 1.1, 16 June 2020)

Abstract

Nakano and Ikeda have proposed a novel indicator called "K" to represent a growth status in COVID-19 spread. The K indicator was proposed for predicting the convergence of the epidemic and early detecting of phase changes in the trend of infection. They not only proposed the simple indicator itself, but also presented a background model on the temporal change of the K value against COVID-19 epidemic.

In this paper, we analyzed the mathematical properties of the K indicator with their temporal change model, and proved that the model is essentially identical to the Gompertz curve which is described by a double-exponential function. In addition, we showed the relationships between the K indicator and other closely related indicators such as "growth rate", "logarithm of the growth rate", and "doubling time". Nakano and Ikeda derived, by fitting the real data, an empirical equation $k = 1 + 2.88K'$, which binds the internal model parameter k with the time derivative (K') of the K indicator. Here we derive the same equation analytically.

The purpose of this paper is not to follow and evaluate the whole work by Nakano and Ikeda, but to clarify mathematical nature of the K indicator. Thus, their excellent applications, i.e. investigation of the differences in epidemic trends between countries based on the decay pace of the K indicator, and the estimation of the number of independent waves of the epidemic in the United States and Japan, for example, are out of scope of this paper. This manuscript is a direct English translation from our previous work, but some notes are newly added on the relationship with the Gompertz curve.

1. Definition of the K indicator

Nakano and Ikeda [1] have proposed a novel indicator called "K" to represent a growth status in COVID-19 spread. The K indicator was proposed for predicting the convergence of the epidemic and early detecting of phase changes in the trend of infection. They not only proposed the simple indicator itself, but also presented a background model on the temporal change of the K value against COVID-19 epidemic.

In this section, the definition of the K indicator is introduced based on the paper by Nakano and Ikeda [1]. A date t represents the number of elapsed days since the predefined reference date, and **the cumulative number of people infected from the reference date to a specific date is noted as $\boldsymbol{N(t)}$.** Then, $N(t)/N(t-7)$, the growth rate on the cumulative number of people infected at one week ago to that on a specific date is focused on. (Note: notation d was used in the original paper [1] instead of t).

By focusing on the growth rate rather than the actual number of infected people, it is easier to grasp the characteristics of the epidemic, which is inherently exponential in nature, and allows for direct comparisons of numbers for countries with different testing regimes and populations. This in itself is not a new proposal, but it is an important starting point. They argued that comparison with a seven-day interval is effective in avoiding distortions in the number of reports that depend on differences of the day of the week, but additional discussion may be required whether this is the optimal setting for the length of the observation interval to capture the essence of the epidemic phenomenon (though it is not treated in this paper).

In this paper, let $\boldsymbol{R(t)}$ **as the growth rate** of cumulative number of infected people with a seven-day interval. (Note: $1 \le R(t)$ because $N(t)$ is cumulative).

$$R(t) = \frac{N(t)}{N(t-7)}, \quad \text{where } t > 7 \tag{1}$$

The K indicator, $K(t)$ of a day t, is defined as follows, as a novel indicator for changes in the cumulative number of people infected. (Note: $0 \le K(t) < 1$).

$$K(t) = 1 - \frac{1}{R(t)} \tag{2}$$

When the infection converges, either temporarily or finally, increase of the cumulative number of infected people gradually slows down, with $R(t)$ approaching 1, and thus $K(t)$ approaching 0.

Nakano and Ikeda dealt with the decreasing of $K(t)$ from about 0.90 to 0.25. With considering the corresponding doubling times (see Section 3) for these indicator values, we can see that the indicator covered a sufficiently wide range of period, from the time of explosive spread of the infection to the time when convergence begins to appear.

To discuss the nature of the K indicator, it is quite informative to consider **the logarithm of the growth rate, $\boldsymbol{L(t)}$**, while Nakano and Ikeda have not mentioned about logarithms in their paper. (Note: $0 \le L(t)$ because $1 \le R(t)$)

$$L(t) = log_e R(t) \tag{3}$$

By using $L(t)$, the definition formula of K indicator (2) can be rewritten as follow.

$$K(t) = 1 - e^{-L(t)} \quad (4)$$

As it is discussed in detail in Section 5, the essence of the proposed model is to assume that the K indicator decays with day t according to a double exponential function (or a Gompertz curve), where k is a constant with $0 < k < 1$ as shown in equation (5) below.

$$K(t) = 1 - e^{-L(0) \cdot e^{-(1-k)t}} \quad (5)$$

where $L(0)$ seems to be not able to calculate from equation (1) and (3) because they are defined for $t > 7$, but $L(0)$ can be easily estimated later.

While it is true that in some range the decrease can be regarded as approximately linear, it is also obvious that not all ranges should be considered linearly.

2. Advantages of K indicator: a simple indicator of decay towards convergence

The superiority of the proposed K indicator is that it can be easily calculated as shown in equation (2), yet it is a value that approaches zero at the convergence of the epidemic with almost monotonic decrease. It is easy to understand intuitively. The value range is between 0 and 1, and that approaching 0 simply means convergence.

A further important feature is empirically shown from the COVID-19 data observed around the world that an almost linear decrease with the progression of date t in a range of the value from $K = 0.9$ to $K = 0.25$. The discovery of this linearity is the major contribution of their paper.

A decrease in $K(t)$ is a decrease in $R(t)$, as can be seen from the definition (2). A decrease in $R(t)$ implies only a slowing of the weekly growth rate of cumulative patient numbers, and the new patient numbers themselves may still be increasing by a large multiple. Put differently, the K indicator is designed to sensitively detect a slowdown in the growth rate itself, even when the number of new patients is still growing significantly.

In their paper, the slope of the temporal change of the K indicator (i.e. time derivative of it) is called K'. K' is the amount of change in the K indicator per day when the K indicator is changing linearly and is usually negative with the decay process. Similar K' values may be observed between culturally similar countries, however, there are several cases where K' values differ significantly, and it is highly demanded to unravel the social, medical, immunological, and/or other reasons for these differences.

Once the slope K' is calculated, it is at first glance possible to extrapolate linearly to find the date t_x with $K(t) = 0$, which means the convergence of the epidemic. However, as we will discuss in Section 4, it is inappropriate to simply use linear extrapolation to find t_x. The paper also suggested a more appropriate method of calculation.

3. Disadvantages of K indicator: difficult to imagine the increase directly

The doubling time (DT) is a closely related indicator to the growth rate $R(t)$. The doubling time is the time it takes for a number to double in an exponential phenomenon. When the exponential parameter representing the phenomenon is constant, the doubling time is also constant.

If we observe the growth rate over the p-day period from date $(t-p)$ to date t, we can calculate **the doubling time $DT(t)$** as follows. The doubling time has a dimension of time and the unit in this paper is a day.

$$DT(t) = \frac{p}{log_2(N(t)/N(t-p))} \tag{6}$$

If the period of observation p is set to 7 days, the equation is as follows, and the close relationship with each of $R(t)$, $L(t)$ and $K(t)$ can be expressed.

$$DT(t) = \frac{7}{log_2(N(t)/N(t-7))} \tag{7}$$

$$= \frac{7}{log_2 R(t)} \tag{8}$$

$$= \frac{7\, log_e 2}{L(t)} \tag{9}$$

$$= \frac{-7}{log_2(1-K(t))} \tag{10}$$

Table 1 shows the relationship between the growth rate $R(t)$, the logarithm of the growth rate $L(t)$, the ratio of R/L, the doubling time $DT(t)$, and K indicator $K(t)$.

It is worthwhile to look carefully at the correspondence between the K indicator and doubling time. For example, $K(t) = 0.90$ corresponds to $DT(t) = 2.11$, which means that the cumulative number of people infected doubled in about two days. On the other hand, $K(t) = 0.25$ corresponds to $DT(t) = 16.87$, which means that the cumulative number of people infected doubled in about half a month. Though this is a relatively slow pace, the same number of new infections appear in the next half month as the cumulative number over the past several months. Even if it decays to $K(t) = 0.15$, which corresponds to $DT(t) = 29.86$, the same number of new infections appear in the next one month as the cumulative number up to the present. The doubling time has been widely used as an intuitive indicator of the status of increase which can directly grasp and estimate the real number of infected people in the coming period.

The actual number of cumulative infections $N(t)$ in the period just prior to the convergence is large, and thus the burden on health care institutions from the new increase is severe, even if the growth rate itself seems to be moderate. The K indicator is attractive, but one small drawback may be that it is difficult to see directly the heavy fact that intense increases continue even with small values of $K(t)$. This issue is further emphasized with the careless use of coarse linear extrapolation, which will be discussed in Section 4.

Table 1 The relationship between the growth rate R, the logarithm of the growth rate L, the R/L ratio, the doubling time DT, and the K indicator K.

days	growth rate	log of the growth rate	R / L ratio	doubling time [day]	K indicator
P	R	$L = Log_e R$	R / L	$DT = P / Log_2 R$	$K = 1 - 1 / R$
7	1.05	0.05	20.52	94.59	0.05
7	1.11	0.11	10.55	46.05	0.10
7	1.18	0.16	7.24	29.86	0.15
7	1.25	0.22	5.60	21.74	0.20
7	1.33	0.29	4.63	16.87	0.25
7	1.43	0.36	4.01	13.60	0.30
7	1.54	0.43	3.57	11.26	0.35
7	1.67	0.51	3.26	9.50	0.40
7	1.82	0.60	3.04	8.12	0.45
7	2.00	0.69	2.89	7.00	0.50
7	2.22	0.80	2.78	6.08	0.55
7	2.50	0.92	2.73	5.30	0.60
7	2.86	1.05	2.72	4.62	0.65
7	3.33	1.20	2.77	4.03	0.70
7	4.00	1.39	2.89	3.50	0.75
7	5.00	1.61	3.11	3.01	0.80
7	6.67	1.90	3.51	2.56	0.85
7	10.00	2.30	4.34	2.11	0.90
7	20.00	3.00	6.68	1.62	0.95
7	33.33	3.51	9.51	1.38	0.97

4. In case the K indicator is assumed to decrease linearly

It is quite interesting that the K indicators calculated from the cumulative number of COVID-19 cases in several countries showed approximately linear decrease with respect to date t over a relatively long period of time. This is one of the benefits of the K indicator. In this section, we examine the requirements of the model by making a false assumption that the K indicator decreases linearly over time.

If the time derivative of $K(t)$ is constant and is represented by the constant K', the following equations hold.

$$K' = \frac{dK}{dt} = \frac{dK}{dR} \cdot \frac{dR}{dt} = \frac{1}{R^2} \cdot \frac{dR}{dt} \tag{11}$$

$$\int K'dt = \int \frac{1}{R^2} dR + C \tag{12}$$

$$K't = \frac{-1}{R(t)} + \frac{1}{R(0)} \tag{13}$$

$$R(t) = \frac{R(0)}{1 - R(0)K't} \tag{14}$$

$$K(t) = \left(1 - \frac{1}{R(0)}\right) + K't \tag{15}$$

As shown on equation (14), the denominator of the $R(t)$ increases by a constant $-R(0)K'$ (Note: $K' < 0$) for each day, and $R(t)$ decreases relatively rapidly. The K indicator $K(t)$ is expressed as a linear function of the slope K' and the intercept $\left(1 - \frac{1}{R(0)}\right)$ as shown in equation (15).

Then the date t_x with $K(t_x) = 0$ is found by the following formula.

$$t_x = -\frac{1}{K'}\left(1 - \frac{1}{R(0)}\right) \tag{16}$$

On this simple model, $K(t) = 0$ and $R(t) = 1$ at date t_x, represented by equation (16), and the cumulative number of people infected $N(t)$ stops increasing completely.

However, the trajectories of the measured K indicators for China and the USA in Figure 1 of the paper [1] showed that the K indicator ceased to fall from the middle of the trajectory and it is revealed that the assumption is unrealistic to go to zero in a straight line.

Here, we explicitly mention the serious caveat so as not to misunderstand their proposal of the K indicator. Although both "the K indicator decreases linearly in some period" and "the various properties of the epidemic can be analyzed from the slope K' in such period" are considered true, it is inappropriate to immediately believe in linear equation (15) in all ranges of time and calculate the convergence date t_x by equation (16). The paper did not propose such calculation.

Their paper proposed an alternative mathematical model for discussing the end stage of the epidemic rather than linear equations (15), and proposed to make observations of the slope K' just to estimate the parameter k of their background model. However, equations (15) and (16) are so clear, and thus there is a danger of misunderstanding.

Verification with the three examples of simulated calculations presented in the Supplement of their paper [1] indicated that $K(t_x) = 0.2$ still remains in actual at the date t_x obtained by equation (16). As shown in Table 1, $K(t) = 0.2$ corresponds to $R(t) = 1.25$, which means that the cumulative number of people infected is still increasing by a factor of 1.25 each week. The number will still increase significantly while the rate continues for some time.

5. The model assuming that the logarithm of the growth rate decreases with time approximately according to a geometric series

In the paper by Nakano and Ikeda [1], it was stated that they found the general property that the K values decay approximately linearly in certain period, and it is also argued that the slope K' of the line is an important parameter, but it is by no means simply believed that the linear equation (15) described in the previous section and that the epidemic converges completely on the date t_x, which is obtained by equation (16).

The model assumed by the paper is more complex and is as follows.

$$N(t+1) = N(t) \cdot e^{a(t)} \tag{17}$$

$$a(t) = k \cdot a(t-1) \tag{18}$$

The cumulative number of infected people $N(\mathrm{t})$ increases monotonically. When the multiplier of the daily increase is expressed as an exponential function, the exponent $a(t)$ changes daily (without loss of generality thus far).

They assumed that the exponent $a(t)$ decays with t as the constant $k (0 < k < 1)$ is multiplied daily as shown in (18). This k is assumed to be a constant for a reasonable period of time (depending on some background of social interventions, cultural lifestyle, and innate immunity of the population, etc.). Needless to reiterate, it is important to remember that the number of $N(t)$ itself will continue to increase, even if the exponent

$a(t)$ decays from day to day, since $a(t)$ is still a positive number.

In order to understand the above model, it is useful to consider the change in $L(t)$ in equation (3). It should be noted that equation (18) above expresses the change from one day to the next, while $R(t)$ in equation (1) and $L(t)$ in equation (3) look at the change from 7 days ago.

$$\begin{aligned} L(t) &= log_e\left(\frac{N(t)}{N(t-1)}\cdot\frac{N(t-1)}{N(t-2)}\cdot\frac{N(t-2)}{N(t-3)}\cdot\cdots\cdot\frac{N(t-6)}{N(t-7)}\right) \\ &= a(t-1)+a(t-2)+a(t-3)+\cdots+a(t-7) \\ &= (k^6+k^5+k^4+k^3+k^2+k+1)\cdot a(t-7) \\ &= \frac{k^7-1}{k-1}\cdot a(t-7) \end{aligned} \tag{19}$$

$$\begin{aligned} L(t+1)-L(t) &= \{a(t)+\cdots+a(t-6)\}-\{a(t-1)+\cdots+a(t-7)\} \\ &= a(t)-a(t-7) \\ &= (k^7-1)\cdot a(t-7) \end{aligned} \tag{20}$$

$$\frac{L(t+1)-L(t)}{L(t)} = -(1-k) \tag{21}$$

According to equation (21), the logarithm $L(t)$ decreases at a rate of $(1-k)$ every day, approaching 0.

From the above discussion, difference and differential equations for $L(t)$ can be formulated as follows, and its solution is obtained.

$$L(t+\Delta t)-L(t) \approx -(1-k)\ L(t)\,\Delta t \tag{22}$$

$$\lim_{\Delta t\to 0}\frac{L(t+\Delta t)-L(t)}{\Delta t} = -(1-k)\ L(t) \tag{23}$$

$$\frac{dL}{dt} = -(1-k)\,L \tag{24}$$

$$L(t) = L(0)\cdot e^{-(1-k)t} \tag{25}$$

where $L(0)$ cannot be directly defined from equation (3), but is easily estimated from $L(7)$ and other data series based on equation (25).

By substituting the obtained equation (25) into equation (4), we obtain equation (5) shown in Section 1.

$$K(t) = 1 - e^{-L(0)\cdot e^{-(1-k)t}} \tag{5}$$

The difference formula (22) is based on a linear approximation to estimate $L(t+\Delta t)$, and it might be more precise, by considering the nature of geometric series, to use the following difference formula (26) instead, and then the solution is obtained as follows.

$$L(t+\Delta t) - L(t) \approx -(1-k^{\Delta t})\ L(t) \tag{26}$$

$$\lim_{\Delta t\to 0}\frac{L(t+\Delta t)-L(t)}{\Delta t} = \lim_{\Delta t\to 0}\frac{-(1-k^{\Delta t})}{\Delta t}\ L(t) = \lim_{\Delta t\to 0}\frac{-(1-k^{\Delta t})\,'}{\Delta t'}\ L(t) \tag{27}$$

$$\frac{dL}{dt} = log_e\, k \cdot L \tag{28}$$

$$L(t) = L(0)\cdot k^t \tag{29}$$

$$K(t) = 1 - e^{-L(0)\cdot k^t} \tag{30}$$

Equation (30) is considered to be a more accurate mathematical formula to express the characteristics of the K indicator model, by its definition. Here the logarithm of the growth rate $L(t)$, for discrete value t, is following a geometric series of the common ratio k as like $a(t)$, The common ratio is k=0.89~0.95 according to the paper [1], thus the logarithm $L(t)$ decreases by 5%-10% on a daily basis.

On the other hand, the equation (5) is still attractive for discussing the overall characteristics of the $K(t)$ curve. **The double-exponential function is also called the Gompertz curve**, which is named after Benjamin Gompertz, a British mathematician. The S-shaped curve has been widely used for modeling animal population saturation, new product dissemination, tumor growth, and so on [2]. (We did not at first realize the relationship between the double-exponential function in equation (5) and the Gompertz curve, but we understand it through a twitter comment [4] given on our previous work about the mathematical analysis of K indicator [3].)

The two curves represented by equation (5) and (30) are only approximately overlaps, but are practically similar enough, because $e^{-(1-k)} \cong k$ holds when $(1-k) \ll 1$, and the common ratio k is almost near 1.0 with using a day as the unit of time.

In the rest of this paper, we employ the double-exponential equation (5) as an approximated curve for the K indicator epidemic model.

Now, based on the above-mentioned double-exponential model, we again consider the time derivative of the K indicator.

$$\begin{aligned}\frac{dK}{dt} &= \frac{dK}{dR}\cdot\frac{dR}{dt}\\ &= \frac{1}{R^2}\cdot\frac{dR}{dt}\\ &= \frac{1}{R^2}\cdot R\frac{d(log_e R)}{dt}\\ &= \frac{1}{R}\cdot\frac{dL}{dt}\\ &= \frac{k-1}{R/L}\end{aligned} \tag{31}$$

As shown in equation (31), the time derivative of the K indicator is not constant and **the R/L ratio appears in the denominator.**

Table 1 shows that the R/L ratio takes a minimum value around $K(t) = 0.65$ and rises slowly from there. The minimum of the R/L ratio can be determined accurately by solving the following equation:

$$\frac{d}{dR}\frac{R}{log_e R} = 0 \tag{32}$$

It is obtained the minimum value of $R/L = e = 2.7183$ when $R = e = 2.7183$ (Napier's number). It corresponds to $K = 1 - \frac{1}{e} = 0.6323$, and then equation (31) takes its maximum.

On the other hand, the R/L ratio when the growth rate over seven days is exactly twice ($R = 2$ and $K = 0.5$) can be obtained by the following.

$$R/L = \frac{2}{log_e 2} = 2.88539\cdots \tag{33}$$

If this fixed R/L ratio can be used to approximate equation (31) around K=0.5, we finally obtain:

$$K' = \frac{dK}{dt} = \frac{k-1}{R/L} \fallingdotseq \frac{k-1}{2.885} \tag{34}$$

$$k \fallingdotseq 1 + 2.885\, K' \tag{35}$$

which is the same equation as the slope of a straight line obtained by the real data fitting in Nakano and Ikeda [1].

6. Conclusion

In this paper, we discussed the mathematical nature of the K indicator proposed by Nakano and Ikeda [1] for detecting the change in the cumulative number of infected people of COVID-19.

The advantages of the K indicator are that it is an intuitive indicator that decays toward convergence and is easily computed by anyone, and that the actual data on COVID-19 in each country suggested that the K indicator decays almost linearly with time.

A minor drawback of the K indicator is that the increasing speed is not directly perceptible, unlike the growth rate R, or doubling time DT. Particular attention is required at times when the epidemic is nearing its convergence and the cumulative number of infected people N becomes large.

The essence of the epidemic model proposed in their work is that the logarithm L of the growth rate decreases towards zero almost following a geometric series as shown in equation (25) or (29). The K indicator itself decreases approximately according to a double exponential function shown in equation (5) or (30).

With following the model as described in Section 5, the time derivative of the K indicator is not inherently constant and depends on the R/L ratio, as shown in equation (31). However, the linear model in equations (34) and (35) is obtained by fixing the R/L ratio at $R/L = 2.885$, which is equivalent to $R = 2$ and $K = 0.5$. Thus, the linear approximation only holds around $K = 0.5$. It is possible to determine to what extent this approximation holds by how much the R/L ratio deviates from the above value.

The K indicator is attractive, and the background model in which the logarithm L decreases following a geometric series is well-fitting to the real COVID-19 data. However, a more accurate approach to finding the common ratio k of the geometric series should be to directly deal with the temporal change in the data of the logarithm L itself, which can be easily calculated, rather than a fit mediated by the approximated equation (35).

References

[1]. Nakano T, Ikeda Y. Novel indicator of change in COVID-19 spread status. medRxiv 2020.04.25.20080200; doi: https://doi.org/10.1101/2020.04.25.20080200

[2]. "Gompertz function" (English version). Wikipedia: The Free Encyclopedia. https://en.wikipedia.org/wiki/Gompertz_function (last accessed 16 June 2020)

[3]. Akiyama Y. Ruiseki-kansensha-su no henka ni taisuru aratana shihyou "K"-chi no riten to ketten ni tsuite (in Japanese) (Advantages and disadvantages of the K indicator, a novel indicator on changes of cumulative number of infected people), Akiyama Laboratory web site, Tokyo Institute of Technology. 10 May 2020. http://www.bi.cs.titech.ac.jp/COVID-19/Mathematical_analysis_of_K_indicator.html (last accessed 16 June 2020)

[4]. petenshicom, [Akiyama has proved that K=1-(double exponential function), but this double exponential function is exactly the Gompertz growth curve...]. Twitter post. 12 June 2020. https://mobile.twitter.com/petenshicom/status/1271361381359513603 (last accessed 16 June 2020)