

累積感染者数の変化に関する新たな指標「K 値」の利点と欠点について

(2020 年 5 月 10 日 東京工業大学 秋山泰)

大阪大学の中野貴志氏と九州大学の池田陽一氏による下記の共著論文（以下、論文）が、医学系プレプリントサーバ medRxiv 上に発表された。拙稿では、同論文において流行の収束時期の予測や感染動向の変化を発見することを目的として提案された新たな指標「K 値」(K indicator)の性質を概観し、その利点と欠点に関する考察を述べる。また、K 値と密接に関連する量である「増加の倍率」「増加の倍率の対数」「倍化時間」等との対応関係に触れるとともに、論文ではデータへの当てはめによって経験的に導出されていた $k = 1 + 2.88K'$ という関係式を、本稿では解析的に導出する。

Nakano T, Ikeda Y. Novel indicator of change in COVID-19 spread status.

medRxiv. 5 May 2020. <https://doi.org/10.1101/2020.04.25.20080200>

本稿の目的は、論文の内容全体を忠実に紹介することではなく、K 値の性質に関して若干の論考を加えることにある。論文ではK値の提案だけに留まらず、K 値の減衰ペースに基づいて世界各国の流行傾向の差異に言及したり、米国と日本における流行の波の数を推定したりもしているが、本稿ではこれらの優れた具体的応用については扱わない。K 値の定義を理解する上で重要となる図表も再掲しないので、必ず原論文を併せて参照されたい。

1. K 値の定義

ある日付を基準日からの経過日数 t で表し、基準日からその日までの累積感染者数を $N(t)$ とおく。同論文では、 $N(t)/N(t-7)$ 、すなわち 1 週間前の累積感染者数から当該日の累積感染者数までの増加の倍率に注目する。(※論文は変数 d を用いていたが、本稿は t を用いる)

実際の人数ではなく、増加の倍率に注目することにより、指数関数的な性質を内在する流行の本質にアプローチしやすいことに加えて、検査体制や人口が異なる各国の数字を直接比較することが可能になる。これ自体は新しい提案ではないが重要な出発点である。一方、必ず 7 日前と比較するという点は、曜日の差異に依存した報告数の歪みを軽減する効果が述べられているが、現象の本質を捉えるための観察区間の長さとして最適な設定であるか否かについて、若干の議論が必要ではないかと思われる（本稿では扱わない）。

本稿では、増加の倍率を $R(t)$ と書く。($N(t)$ は累積数であるため、 $1 \leq R(t)$)

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

論文では、以下で計算される値 $K(t)$ を、日付 t における「K 値」(K indicator)と名付け、累積感染者数の変化に関する新たな指標とすることが提案された。($0 \leq K(t) < 1$)

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

感染が一時的または長期的に収束するとき、累積感染者数の増加は徐々に鈍化し、 $R(t)$ は1に近づき、 $K(t)$ は0に近づく。論文では、 $K(t)$ の値が0.90程度から0.25程度に減少するまでの領域を特に詳しく取り扱っているが、これらの値に対応する倍化時間（第3節）を考えてみると、爆発的に感染が広がる時期から、収束が見え始める時期までを、十分に広くカバーできていることがわかる。

論文では触れられていないが、 K 値の性質を論じる上で、**増加の倍率の対数**である $L(t)$ を考えてみるのがきわめて有益である。（ $1 \leq R(t)$ であるため、 $0 \leq L(t)$ ）

$$L(t) = \log_e R(t) \quad (3)$$

本稿で導入した $L(t)$ を用いて、 K 値の定義である(2)式を書き直すと以下のようになる。

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

第5節で詳しく述べるように、提案されたモデルの本質は、日付 t にしたがって K 値が以下の(5)式のような二重指数関数（ k は、 $0 < k < 1$ である定数）に従って減衰すると仮定するものである。ある区間では近似的には直線的に減少するというのは事実であるが、すべての区間を線形で考えてはいけないことも明白である。

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

2. K 値の利点：収束に向かって減衰する分かりやすい指標

提案された K 値が指標として優れている点は、(2)式のように誰にでも簡便に計算できる数値であるのに、ほぼ単調に減少しながら感染の収束時に 0 に近づく量であることである。値域が 0～1 で、0 に近づけば収束を意味するという性質は、直観的に理解しやすい。

さらに重要な特徴として、おおよそ $K = 0.9$ から $K = 0.25$ までの範囲においては、日付 t の進行に伴ってほぼ直線的に減少することが、世界各国の観測値から経験的に示されている。この線形性を発見したことが、当該論文の大きな貢献である。

$K(t)$ が減少することは、定義からわかるように $R(t)$ が減少することである。 $R(t)$ の減少は、累積患者数の週当たりの増加倍率が鈍ることだけを意味しており、新規患者数の数字自体は、まだ大きな倍率で増加を続けているかもしれない。逆の言い方をすれば、新規患者数が大きく増大している状況であっても、増加倍率の鈍化そのものを鋭敏に察知できるように、 K 値は設計されている。

論文では、 K 値の時間的変化の傾きを K' と呼ぶ。 K' は、 K 値が直線的に変化しているときの 1 日あたりの K 値の変化量で、通常は減衰過程を扱うため負の値をとる。文化的に似た国々の間で類似した K' 値が観測される場合もあるが、 K' が大きく異なる場合もあり、これらの差異が生じる社会上・免疫上・医学上などの理由の解明が待たれるところである。

傾き K' が算出できれば、直線を外挿して、 $K(t) = 0$ となる日付 t_x 、すなわち流行の収束の

日付を求めることも一見可能である。ただし、第4節で述べるように、直線を単純に外挿して t_x を求める行為は好ましくない。論文内ではより好適な計算の方法が示唆されている。

3. K 値の欠点：値から直接に増大ペースを想像することが容易ではない

増加の倍率 $R(t)$ と密接に関連した指標として、倍化時間(DT: doubling time)がある。

倍化時間とは、指数関数的な増加を示す現象において、数値が 2 倍になるのに要する時間を表し、現象を支配する指数が一定であるときは倍化時間も一定となる。日付 $(t-p)$ から日付 t までの p 日間の増加の倍率を観測すれば、**倍化時間 $DT(t)$** を以下のように計算できる。倍化時間は時間の次元をもち、本稿における単位は日である。

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

観察の期間 p を 7 日間と定めれば次式となり、 $R(t), L(t)$ および $K(t)$ との関係が理解される。

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

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

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

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

表 1 に、増加の倍率 $R(t)$ 、増加の倍率の対数 $L(t)$ 、 R/L の比率、倍化時間 $DT(t)$ 、および $K(t)$ の対照表を示す。

K 値と倍化時間との対応を注意深く見ておくことには価値がある。 $K(t) = 0.90$ は、 $DT(t) = 2.11$ に相当し、約 2 日ごとに累積感染者数が倍々になるという状況である。一方、 $K(t) = 0.25$ は、 $DT(t) = 16.87$ に相当し、約半月で倍になるという比較的遅いペースである。しかしそれでも、過去数か月の累計感染者数と同数の新規感染者が次の半月で現れる。 $K(t) = 0.15$ まで減衰しても、 $DT(t) = 29.86$ に相当し、次の約 1 か月でそれまでの累積感染者数と同数が新たに発生することを意味する。このように倍化時間は、感染者数の増大のペースそのものを直観的に把握できる指標であり、広く用いられてきた。

収束直前の時期における累積感染者数 $N(t)$ の実際の数値は大きいので、たとえ増加の倍率は穏やかになっていても、新たな増加による医療機関への負荷は深刻である。K 値は魅力的な指標だが、 $K(t)$ が小さな値でも激しい増加は続いているという重大な事実が直接的に見えにくい点は、一つの小さな欠点といえるかもしれない。仮に、第4節で述べる直線の粗い外挿のような不注意な利用と組み合わせられてしまうと、この問題点はさらに強調される。

表1 増加倍率 R 、増加倍率の対数 L 、 R/L 比、倍化時間 DT 、 K 値の対照表

日数	増加倍率	増加倍率 の対数	R/L 比	倍化時間 (日)	K 値
P	R	$L = \text{Log}_e R$	R/L	$DT = P / \text{Log}_2 R$	$K = 1 - 1/R$
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. K 値が直線的に減少すると仮定した場合

各国の COVID-19 の累積感染者数から算出された K 値が、比較的長い区間において、日付 t に対しておおよそ直線的に減少する傾向が見られたことは興味深く、指標としての K 値が持つ魅力の一つになっている。

本節では、いったん K 値が真に線形に減少すると仮定してみることにより、その背後に要求されるモデルが満たすべき性質を検討する。

仮に、 $K(t)$ の時間微分が一定であり、定数 K' で表されるとすると、

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

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

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

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

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

増加の倍率 $R(t)$ は、(14)式のように分母が1日ごとに定数 $(-R(0)K')$ ずつ増加しつづけ
($K' < 0$ に注意)、比較的急速に減少していく。 K 値 $K(t)$ は、(15)式のように切片 $\left(1 - \frac{1}{R(0)}\right)$
の直線の方程式で表される。直線が t 軸と交わり、 $K(t) = 0$ となる日付は次式で求められる。

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

この単純なモデルの上では、(16)式で表される日付 t_x において、 $K(t) = 0$ および $R(t) = 1$ になり、累積感染者数 $N(t)$ の増加は完全に止まる。しかし論文の Figure 1 に掲載されている China や USA の実測された K 値の軌跡を見る限り、 K 値は途中から下げ止まり、直線のままで 0 に向かうことは現実的ではないことが読み取れる。

ここで、論文の提案内容を誤解しないための、極めて重大な注意点を述べたい。「 K 値は一部の区間では直線的に減少すること」および「その区間における傾き K' から流行の諸性質が解析できること」はどちらも事実と考えられるものの、そこから直ちに (15)式の直線の方程式を信じて、(16)式により収束時期 t_x を算出することは不適切である。そのようなことは論文上で提案されていない。

論文では (15)式の直線の方程式ではなく、流行の終盤の時期を議論するための別の数理モデルを提案しており、傾き K' の観測は、あくまでもその数理モデルのパラメータ k を推定するためにだけ使うことを提案している。しかし、(15)式と(16)式があまりにもわかりやしいため、上記のような誤解が生じることが若干危惧される。論文の Supplement で紹介されている 3 つの模擬計算の例を用いて検証してみると、(16)式で求めた日付 t_x では実際には $K(t) = 0.2$ 程度の残存があることが示唆される。前節の対照表に示すとおり、 $K(t) = 0.2$ は $R(t) = 1.25$ に相当し、累積感染者数は毎週 1.25 倍ずつの増加がまだ続いている時期である。そしてそのペースがしばらく続く間に、累積感染者数 $N(t)$ の実数はまだ大きく増大する。

5. 増加倍率の対数が等比数列に従って減少すると仮定した場合

論文においては、 K 値がある区間でおおよそ直線的に減衰する性質を発見したと述べられており、直線の勾配 K' が重要なパラメータであることは論じられているが、決して前節で述べた直線の方程式(15)式を単純に信じて、(16)式で求められる日付 t_x に流行が完全に収束すると考えられているわけではない。

論文が仮定しているモデルは、より複雑であり、以下のようなものである。

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

$$a(t) = k \cdot a(t-1) \quad (18)$$

累積感染者数 $N(t)$ は累積数であるために必ず増加していくが、1日ごとの増加の倍率を指数関数で表現し、その指数 $a(t)$ は日々変化するものとする(ここまでは一般性を失わない)。これ以降は論文が採用したモデル上の仮定となるが、指数 $a(t)$ に対して、定数 $k(0 < k < 1)$ が毎日乗じられて、 t とともに減衰していくものとしている。この定数 k は、(社会的介入や生活様式や国民の自然免疫等の何らかの背景によって決まり)、相応の期間は一定していると仮定されている。改めて述べるまでもないが、指数 $a(t)$ が日々減衰していくとしても、指数関数の引数に正の数 $a(t) > 0$ が入るのであるから、 $N(t)$ の数値自体は増え続けていくことを忘れてはならない。

上記のモデルを理解していく上では、(3)式の $L(t)$ の増減を考えることが有益である。

その際に、上記の(18)式は1日ごとの変化を表す式であるのに対して、(1)式の $R(t)$ や、(3)式の $L(t)$ は7日前との変化を見ている点に、多少注意する必要がある。

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

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

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

(21)式によれば、対数 $L(t)$ は毎日 $(1 - k)$ の割合で減少して0に近づいていく。

すなわち、増加の倍率 $R(t)$ の対数 $L(t)$ が公比 k の等比数列となる。論文内では、 $k =$

0.89~0.95であったので、毎日 5%~10%程度ずつ $L(t)$ の値が減少していくことになる。

上記の議論から、 $L(t)$ に関する微分方程式が立てられ、その解が以下のように求まる。

$$\frac{dL}{dt} = -(1-k)L \quad (22)$$

$$L(t) = L(0) \cdot e^{-(1-k)t} \quad (23)$$

得られた(23)式を、(4)式に代入すれば、第1節で示した(5)式を得る。

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

ここで上記のモデルに基づいて、 K 値の時間微分について、再度考える。

$$\begin{aligned} \frac{dK}{dt} &= \frac{dK}{dR} \cdot \frac{dR}{dt} \\ &= \frac{1}{R^2} \cdot \frac{dR}{dt} \\ &= \frac{1}{R^2} \cdot R \frac{d(\log_e R)}{dt} \\ &= \frac{1}{R} \cdot \frac{dL}{dt} \\ &= \frac{k-1}{R/L} \end{aligned} \quad (24)$$

(24)式に示されたとおり、 K 値の時間微分は一定ではなく、**分母に R/L 比が登場する**。

表1を改めて見ると、 R/L 比は $K(t) = 0.65$ 前後で最小値を取り、そこから離れるにつれて緩やかに上昇することがわかる。 R/L 比の最小値については、正確に求めることができ、

$$\frac{d}{dR} \frac{R}{\log_e R} = 0 \quad (25)$$

を解くことにより、 $R = e = 2.7183$ (ネイピア数) のとき、すなわち $K = 1 - \frac{1}{e} = 0.6323$ のときに、最小値 $R/L = e = 2.7183$ を取る。

一方、7 日間での増加倍率がちょうど 2 倍となる $R = 2$ のとき、すなわち $K = 0.5$ のときの R/L の値は次式で計算される。

$$R/L = \frac{2}{\log_e 2} = 2.88539 \dots \quad (26)$$

仮に、この時の R/L 値を用いて、(24)式を $K = 0.5$ の周辺で近似できると考えるならば、

$$K' = \frac{dK}{dt} = \frac{k-1}{R/L} \approx \frac{k-1}{2.885} \quad (27)$$

$$k \approx 1 + 2.885 K' \quad (28)$$

となり、論文中でデータへのフィッティングで得た勾配 K' ("a slope of a straight line obtained by the fit"、論文より引用) と同じ式となる。

6. まとめ

本稿では、中野貴志氏と池田陽一氏が提案した累積感染者数の変化に関する新たな指標「K 値」の性質について、若干の考察を加えた。

K 値の利点は、誰にでも簡便に計算できる数値でありながら、収束に向かって減衰する直観的に分かりやすい指標であることと、ある区間において、K 値が時間とともにほぼ直線的に減衰することが、COVID-19 に関する各国の実データから示唆されている点である。

K 値の欠点があるとすれば、関連する量である増加の倍率 R や、倍化時間 DT とは異なり、倍増ペースが直接的には感じ取りにくい点であり、特に流行の収束間際で累計感染者数 N が大きくなっている時期には注意を要する。

論文で提案されている流行モデルの本質は、増加の倍率 R の対数 L が、(23)式に示すように等比数列として 0 に向かって減少していくことである。指標である K 値そのものは、(5)式に示すように二重指数関数に従って減少していく。

このモデルに従うならば、K 値の時間微分は、(24)式に示すように本来は一定ではなく R/L 比に依存する。しかしここで、 $R = 2$ のとき、すなわち $K = 0.5$ のときの、 $R/L = 2.885$ の値を用いて R/L 比を固定してしまったのが、(27)式および(28)式で表される直線のモデルである。直線の近似は $K = 0.5$ の周辺でのみ成立するものであり、どの区間までこの近似が成り立つかは、 R/L 比が上記の値からどれくらい乖離するかで判断することが可能である。

K 値は魅力的な指標であり、増加の倍率の対数 L が等比数列として減少するというモデルも現実のデータとフィットしているという点で興味深い。等比数列の公比 k を求める際には、近似された(28)式を仲介しての当てはめではなく、簡単に計算できる増加の倍率の対数 L のデータの変化そのものを扱うことが、より正確なアプローチではないかと考えられる。