Logistic function and complementary Fermi function for simple modelling of the COVID-19 pandemic: tutorial notes

John R Barker
James Watt School of Engineering, University of Glasgow, Glasgow G12 8LT, Scotland, UK

John.Barker@glasgow.ac.uk
28th March, 2020

Abstract: Some tutorial notes on the three parameter logistic function and its relation to the Fermi function for simple modelling pandemics phenomena. The WHO data for the number of cases of infection with the COVID-19 virus for multiple countries as a function of time is fitted very well by a discretised logistic function. The rate of infection may be modelled by the derivative of the logistics function. More complex phenomena such as the appearance of fresh outburst of infection may be treated by using a superposition of logistic functions. The data for China and Italy are analysed. These notes are intended for undergraduates, graduate students who may find pleasure in the ability of applied mathematics and physics to shed light on the complexities of the real world.

1.Introduction

The coronavirus COVID-19 outbreak is now a pandemic. The World Health Organisation is collating and publishing the number of infections, the rate of infection and the number of deaths on a daily basis for each reporting country [1]. The disease was first reported in Wu-Han city in Hubei Province, China (January 2020) and the outbreak there now (28th March 2020) appears to be almost at a standstill. Figure 1 plots the number of reported infections in China for 55 days after the 278 count in early January 2020. Figure 2 shows the rate of infections (increase in cases from previous day). The reported data is in the Appendix.

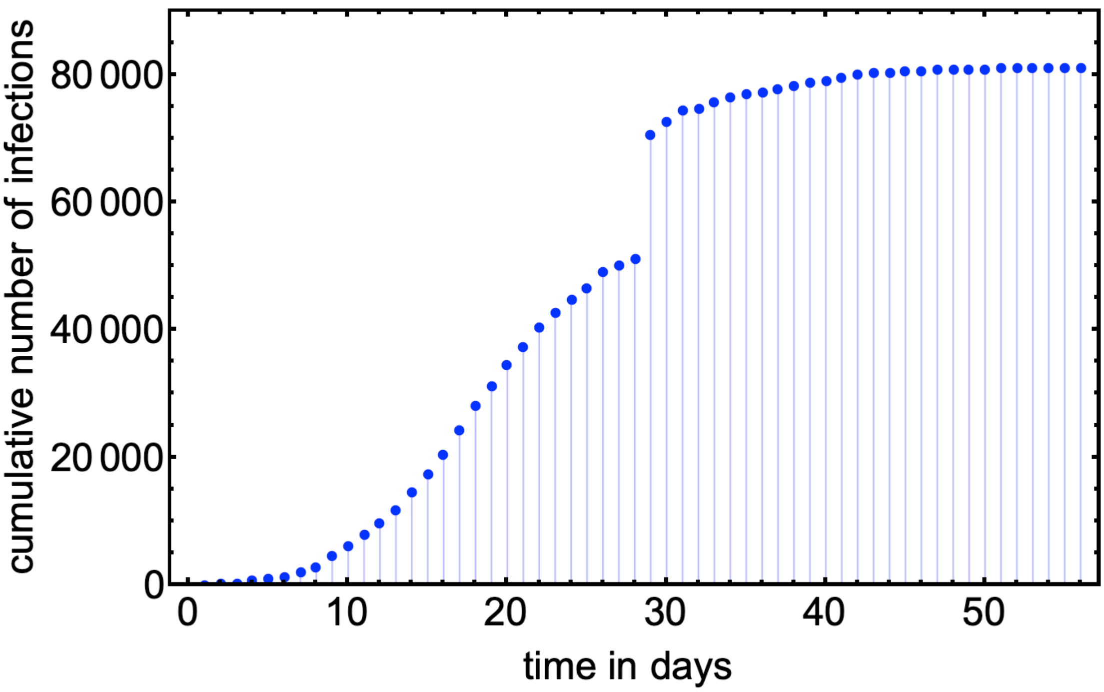

Figure 1: Cumulative infections per day. Data mainly from Wubei area China [1].

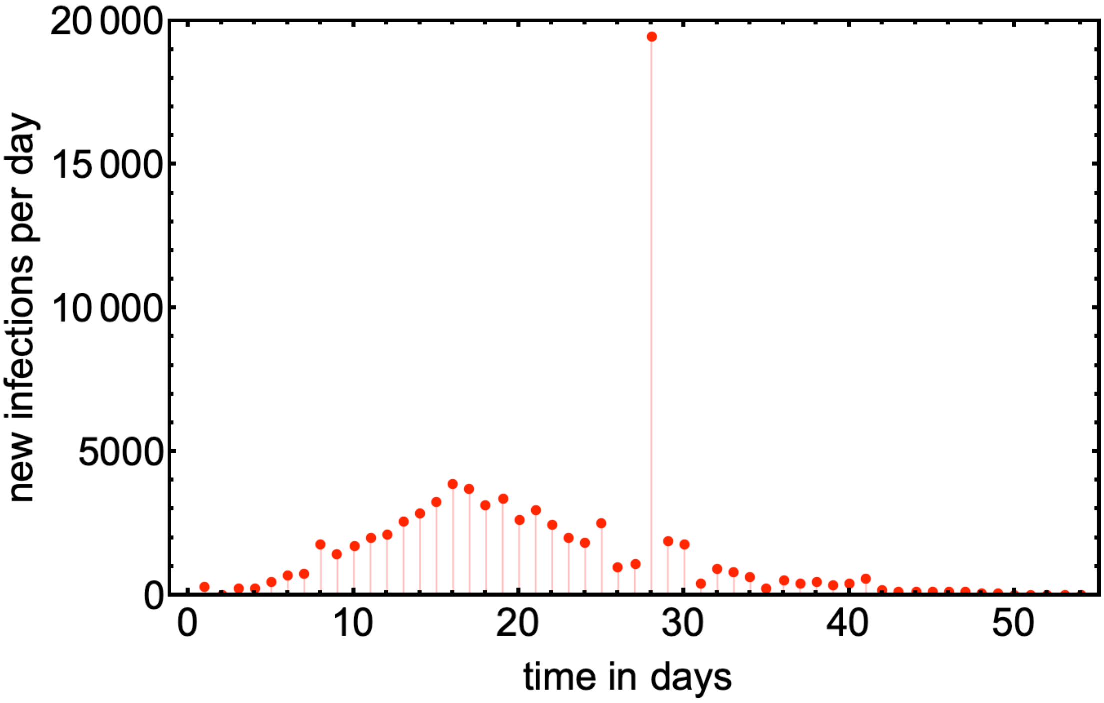

Figure 2: Rate of infection (new cases per day). Data mainly from Wubei area, China [1].

These real data plots are marred by a discontinuity at day 29 *where the counting methodology was corrected*. Such a feature is characteristic of a great deal of the WHO data where different countries have amended their counting/reporting methodology at different times.

Disregarding the discontinuity, the number of infections in Figure 1a is seen to rise with time in an approximately exponential fashion, reaching a maximum infection rate and later levelling off.

For comparison, Figures 3-4 show the first 37 days of the outbreak in Italy [1] displaying the number of infections and the rate of infection (number of new cases per day). Here the Italy data shows the exponential growth of the infections appears to be approaching an inflection point as suggested by a possible peak in the rate of infections.

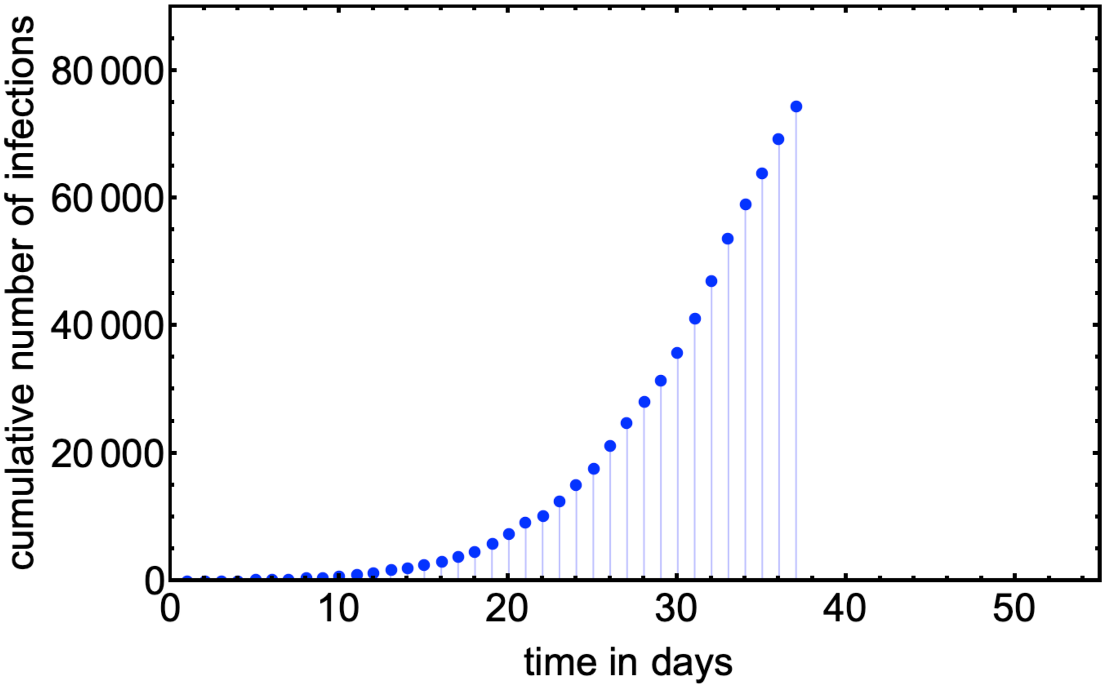

Figure 3: Cumulative infections per day. Data from Italy [1].

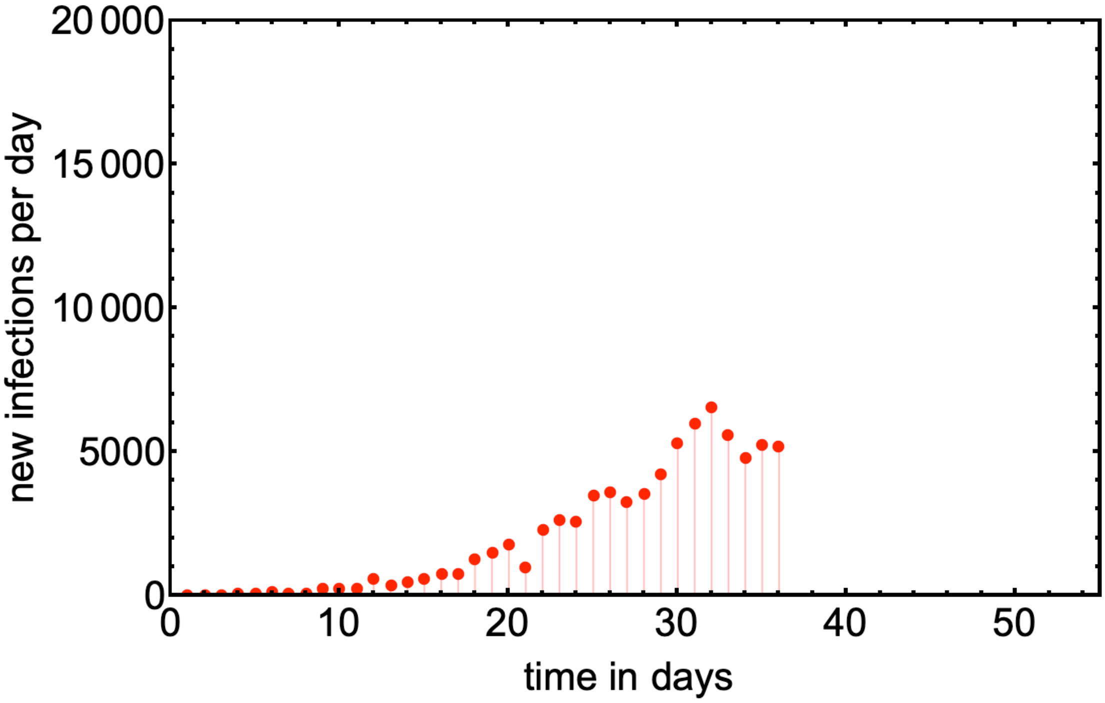

Figure 4: Rate of infection (new cases per day). Data from Italy [1].

These two examples show the real world noisy discrete data that is facing epidemiologists in their attempts to predict the outcome of the pandemic and the consequences of counter-measures. The fully-developed spread and containment of the outbreak as shown in the China data is reminiscent of a "learning curve" and of **2.**

various technological “laws” such as Moore’s Law [2] that describes the annual increase in the maximum number of transistors per integrated circuit due to advances in miniaturisation. With all “learning curves” we observe a sigmoidal shape that starts off slowly, rises to a maximum rate of learning and then levels off [3]. In section 2 we describe a simple three parameter model for this process in terms of the logistic function (introduced by Verhulst [4] in a study of population dynamics) or the related complementary Fermi function [5] widely used in semiconductor physics. The logistic function [6] is just one representation of the step function (actually a distribution) and we have used this recently in discussing the electrostatic self-energy in nanowire transistors [7]. In section 3 we show the application to the COVID-19 pandemic data. In the Appendix we list some of data utilised. All the calculations were carried out in *Mathematica*®.

2. The logistic function

2.1 Basics

For our purposes we write the logistic function (or sigmoid function [6]) as:

$$f(t) = \frac{c_{\max}}{1 + e^{-(t-t_{peak})/T}} \tag{1a}$$

$$= \frac{c_{\max} e^{(t-t_{peak})/T}}{1 + e^{(t-t_{peak})/T}} \tag{1b}$$

Where we might identify $f(t)$ with either the number of infections or the number of deaths. The first derivative of f is the rate (of infection) $r(t)$:

$$r(t) = f' = \frac{\mathrm{d}f}{\mathrm{d}t} = \frac{(c_{\max}/T)e^{-(t-t_{peak})/T}}{(1+e^{-(t-t_{peak})/T})^2} \tag{2a}$$

f satisfies the differential equation:

$$\frac{d(f/c_{\max})}{dt} = \frac{(f/c_{\max})\{1 - f/c_{\max}\}}{T}; \text{boundary condition } f(t_{peak}) = c_{\max}/2 \quad \text{(2b)}$$

Equation 2b has a very simple interpretation: the probability of an individual being infected at time t is $p(t) = f(t)/C_{max}$; the probability of not being infected is therefore $(1\text{-}p)=(1\text{-} f(t)/C_{max})$; so the rate of increase in infections dp/dt is set proportional to the conditional probability $p(1\text{-}p)$ that someone will be infected if they are not already infected.

Figure 5 illustrates the logistic function and its first derivative for the parameters $c_{\max} = 100, t_{peak} = 50, T = 5, \tau = 3.46$. Figure 6 illustrates the first two derivatives of the logistic function and the exponential approximation for the same parameters.

2.2 Interpretation of the parameters: $c_{\max}, T$ *and* t_{peak} .

(i) c_{max} is the asymptotic value of f for $t >> t_{peak}$.

For $t << t_{peak}$: f is asymptotic to 0.

(ii) t_{peak} is the value of t for which df/dt is a maximum (**the peak rate**).

(iii) $f(t_{peak}) = c_{max}/2$ (3)

(iv) For $t << t_{\mathrm{peak}}$:

$$f(t) = \frac{c_{\max} e^{(t-t_{peak})/T}}{1+e^{(t-t_{peak})/T}} \underset{(t-t_{peak})/T<0}{\longrightarrow} c_{\max} e^{(t-t_{peak})/T} = f_{\mathrm{e}} \tag{4a}$$

$f_e(t)$**,** Equation (4) **is the exponential approximation**.

In terms of the natural logarithm (4a) may be written as

$$\begin{aligned} Ln[f_e(t)] &= Ln[c_{\max}] + (t - t_{peak})/T = a + t/T \\ a &= Ln[c_{\max}] - t_{peak}/T = \text{constant} \end{aligned} \tag{4b}$$

Eqn (4b) is a linear equation in t with gradient 1/T. The rate df_e/dt is simply

$$f_e^{'}(t) = \frac{f_e(t)}{T} \tag{5}$$

(v) *T: a decay/rise time* is a measure of the time τ for f to double in value in the exponential approximation:

$$\tau = Ln(2)T = 0.69314T \tag{6}$$

(vi) The discrete logistic function is defined for integer values of t as

$$f_{\mathrm{int}}(t) = Int[f(t)], t \in Integers \tag{7}$$

(vii) The logistic function is the integral of the rate $r(t)$

$$f(t) = \int_{-\infty}^{t} \mathrm{d}t' r(t') \tag{8}$$

It follows from (8) that the area under the rate curve gives c_{max}.

$$c_{\max} = \int_{-\infty}^{\infty} \mathrm{d}t' r(t') \tag{9}$$

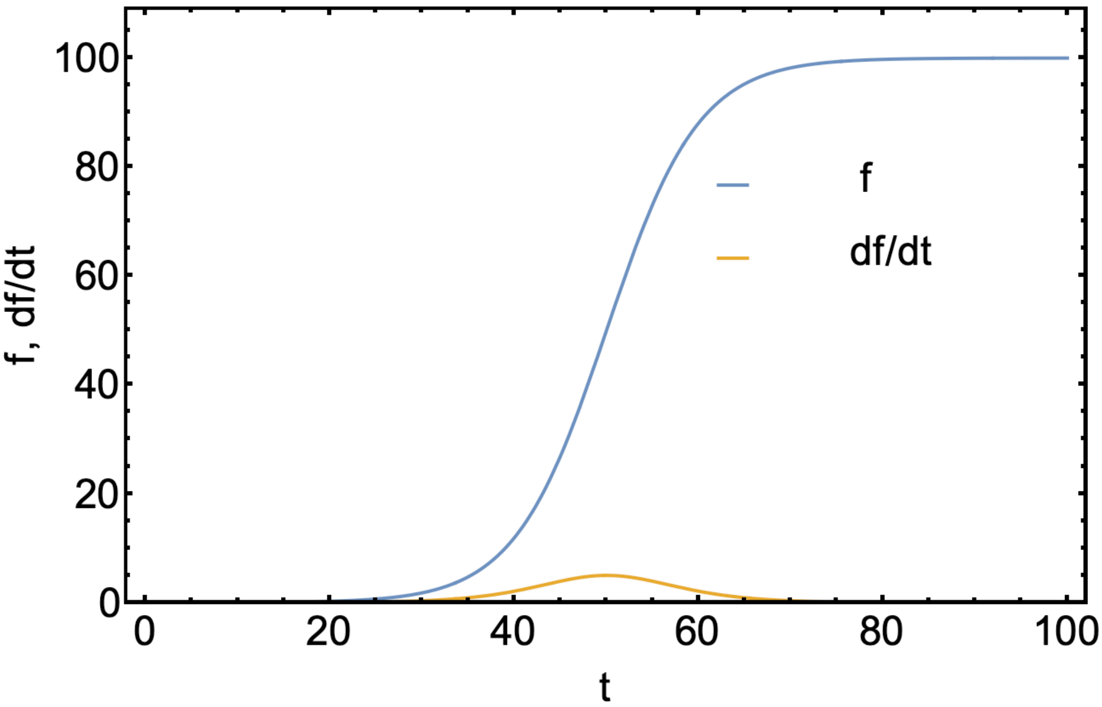

Figure 5: The logistic function $f(t)$ and its first derivative $f'(t)$ for the parameters $c_{\max} = 100, t_{peak} = 50, T = 5, \tau = 3.46$

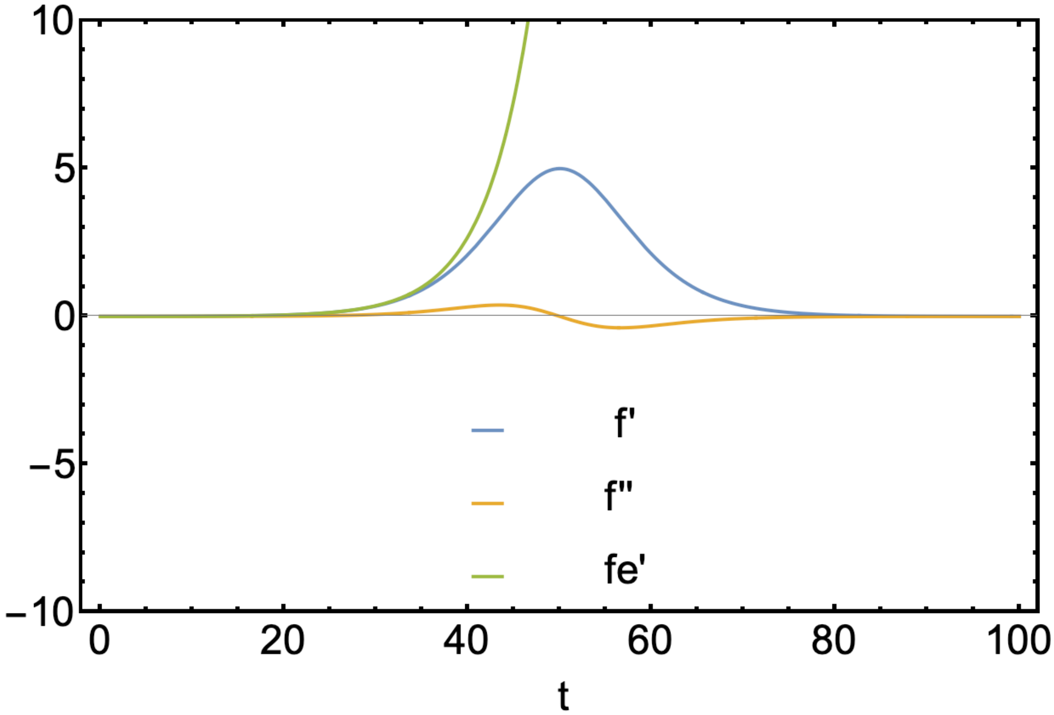

Figure 6: Plots of f', f'', the first and second derivatives of the logistic function and f_e' the first derivative of the exponential approximation for the parameters $c_{\max} = 100, t_{peak} = 50, T = 5, \tau = 3.46$

From (2) and (9) observe that if t_{peak} and/or T are altered , for example, flattening the peak rate or advancing its occurrence, the total number of infections c_{max} is unchanged (see figure 7).

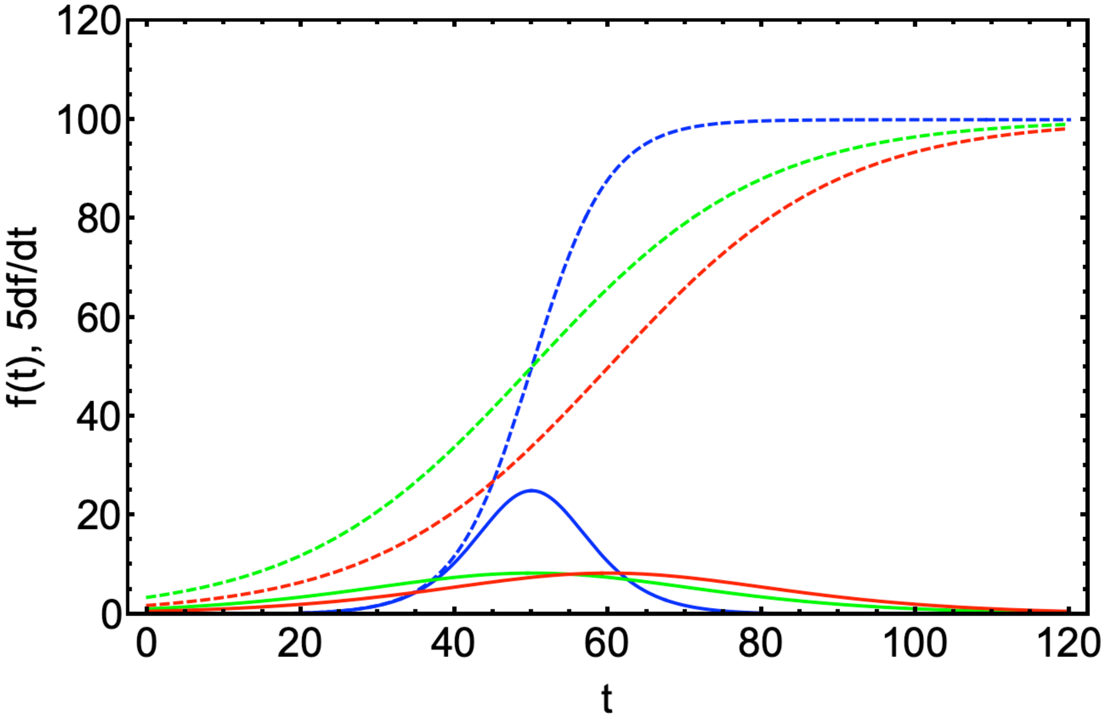

Figure 7: The total number of infections c_{max} is unaffected by altering the values of t_{peak} and/or T.
Blue curves: *f(t)*, 5df/dt, c_{max}=100, t*peak*=50, *T*=5;
Green curves: *f(t)*, 5df/dt, c_{max}=100, t*peak*=50, *T*=15;
Red curves: *f(t)*, 5df/dt, c_{max}=100, t*peak*=60, *T*=15.

2.3 Relation to the Fermi-Dirac function

The logistic function is related to the quantum statistical distribution function known as the Fermi-Dirac function ([5] well-known in physics, particularly condensed matter and semiconductor physics):

$$f_{FD}(\varepsilon;\mu,\beta,1)=\frac{1}{1+e^{\beta(\varepsilon-\mu)}} \tag{10}$$

Here ε is the energy, μ is the Fermi energy, $\beta=1/k_B T$ is the inverse of the product of Boltzmann's constant and the absolute temperature T *(we use* β *so as not to be confused with the decay/rise time T!)*. In terms of the logistic function (written with the parameter dependence made explicit) we find:

$$f\,(t;t_{peak},T,c_{\max})=\frac{c_{\max}}{1+e^{-(t-t_{peak})/T}} \tag{11}$$

$$f_{FD}(\varepsilon;\mu,\beta,1)=1-f\,(\varepsilon;\mu,1/\beta,1) \tag{12}$$

$$f(t;t_{peak},T,c_{\max}) = c_{\max}\{1 - f_{FD}(t;t_{peak},1/T,1)\} \tag{13}$$

Thus the logistic function is essentially a complementary Fermi-Dirac function.

2.4 Alternatives

There are many generalisations to the logistic function, for example, all or some of the fixed parameters may be made to be time-dependent. The general qualitative form of the logistic function may be modelled by other 3-parameter functions, for example, in terms of the error function *erf*[*x*] we may define:

$$f_{error}(t) = \frac{c_{\max}}{2T\sqrt{\pi}}\int_{-\infty}^{t} dt' e^{-(t-t_{peak})^2/4T^2} = c_{\max}(1 + erf[(t - t_{peak})/2T]) \tag{14}$$

$$r_{error}(t) = \frac{\mathrm{d}f_{error}(t)}{\mathrm{d}t} = \frac{c_{\max}}{2T\sqrt{\pi}} e^{-(t-t_{peak})^2/4T^2} \tag{15}$$

A comparison with the logistics function and its first derivative is shown in Figure 8 for the same values of the three parameters. The results are similar but unlike the logistics curves the error function model does not possess an exponential growth region.

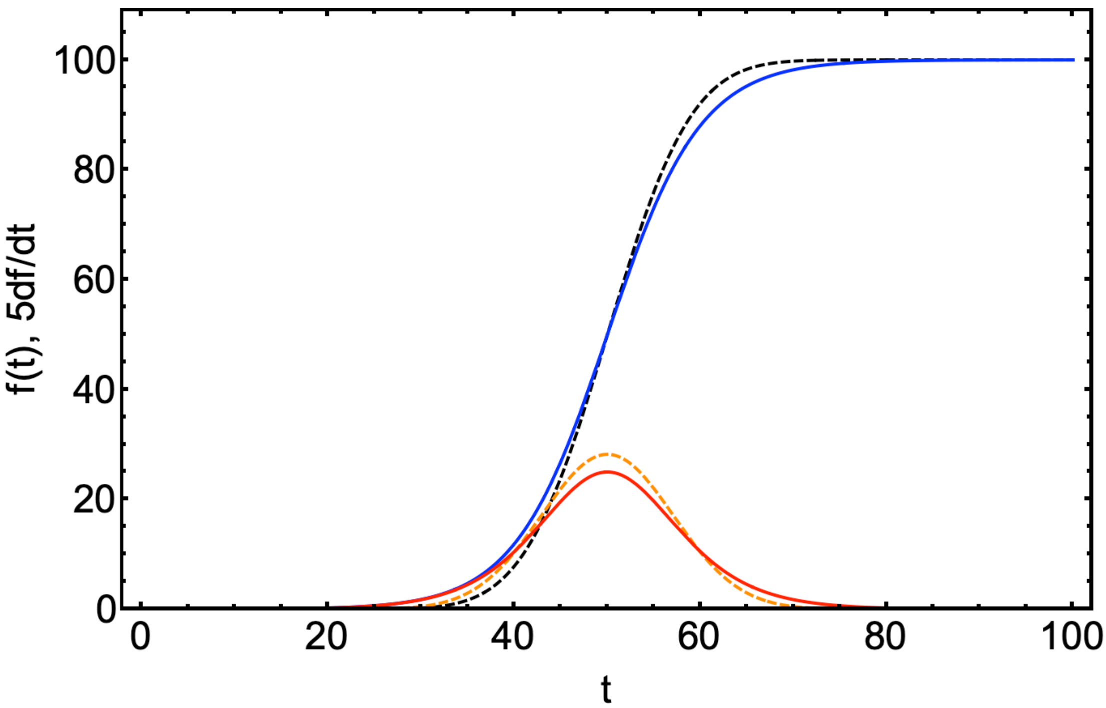

Figure 8: Comparison of the logistics model (full lines) with the error function model (dashed lines) for the same parameters. Upper curves are for the direct functions. The lower two curves are the first derivatives. Scaled by a factor 5. The parameters used are: c_{max}=100, t_{peak}=50, T=5.

3. Simple application to pandemics

3.1 Extracting the parameters

The three parameter logistic function may be easily fitted to the observed cumulative infection data provided the data extends beyond the peak in the infection rate. There are simple procedures for doing this especially using *Mathematica*® which deploys various fitting algorithms. However, in earlier stages of an epidemic or pandemic one may extract the decay/rise time T from the approximate exponential regime of the logistic function and indeed the infection rate (this easily performed by taking the logarithm of the infection as in equation 4b). The noisy nature of actual observations may make this very approximate but using the logarithm of the infection rate there are strong statistical methods for extracting the parameter T. The determination of the other parameters is very difficult in this regime. However, if an inflection in the cumulative infection is observed and/or a peak is observed in the infection rate then t_{peak}, the time of the peak infection rate may be estimated. It follows from equation (3) that the asymptotic cumulative infection c_{max} may estimated as:

$$c_{\max} \approx 2 \times \text{(the observed cumulative infection at the time of the observed peak in the infection rate).} \qquad (16)$$

From these parameters one may estimate (within the limitations of the model) the subsequent cumulative infection and infection rate. A similar procedure holds for cumulative deaths/death rate due to the virus but the three parameters will not generally be the same. The cumulative deaths and the death rate are reasonably objectively based. The true c_{max} parameter for the cumulative infections may be estimated: (i) provided the morbidity data are combined with other epidemiological estimates; (ii) widespread testing provides an estimate of those with mild or no apparent symptoms.

From the point of view of managing a pandemic crisis it is clear from Figure 7 that delaying and/or flattening the infection rate peak (by for instance social isolation or lockdown measures) is of benefit in spreading the hospitalisation load of seriously infected persons, but the asymptotic cumulative infection will remain the same as discussed in section 2.2. However, if more at risk patients can be hospitalised the lower the eventual number of deaths. Preventative measures such as social isolation, tracking, testing and lockdowns may temporarily reduce the true values of c_{max} for the cumulative infection but ultimately c_{max} will be some significant factor of the total population.

3.2 Multiple or delayed outbreaks

Suppose an epidemic (e.g in a single country) or pandemic involves delayed outbreak. For example, let a secondary outbreak occur after a simple outbreak is well developed, then we might postulate an additive model function with 3N parameters

$$f_{total}(t;t_{peak1}...t_{peakN},T_1...T_N,c_{\max 1}...c_{\max N}) = \sum_{I=1}^{N} f_{total}(t;t_{peakI},T_I,c_{\max I}) \qquad (17)$$

Figure 9 illustrates a hypothetical epidemic outbreak with two phases based on equation (17).

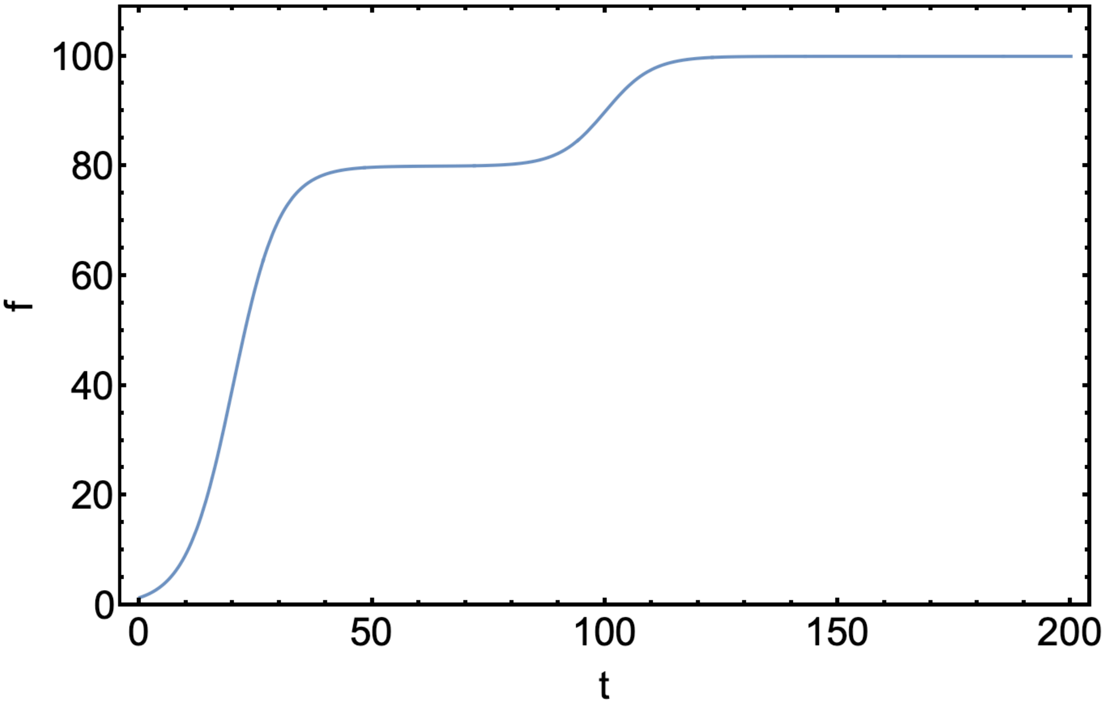

Figure 9: A six parameter multiple logistics function with parameters:

c_{max1} = 80, $tpeak_1$ = 20, T_1 = 5; c_{max2} =20, $tpeak_2$ =100, T_2 = 5

A first outbreak peaks at 20 days is followed by a second outbreak peaking at 100 days.

3.3 Application to the COVID-19

We now show the results of applying the above analysis to the WHO data for the spread of infections in the World, China and Italy.

3.3.1 China data

Figure 10 shows the WHO data (blue dots) for the cumulative infections and daily rate of infection(red dots) for China with superposed plots of a logistic function *f* for the cumulative infections and daily rate= of infections d*f*/dt. The logistic model parameters are c_{max} = 81500 , t_{peak} = 20, T = 5 days. Despite the previously mentioned counting change in the WHO data the logistic model shows a very good fit with an implied peak in the infection rate at 20 days. The level off predicted by the logistic model is not supported by the reported data which is beginning to show small new outbreaks in China mainly due to imported cases according to WHO.

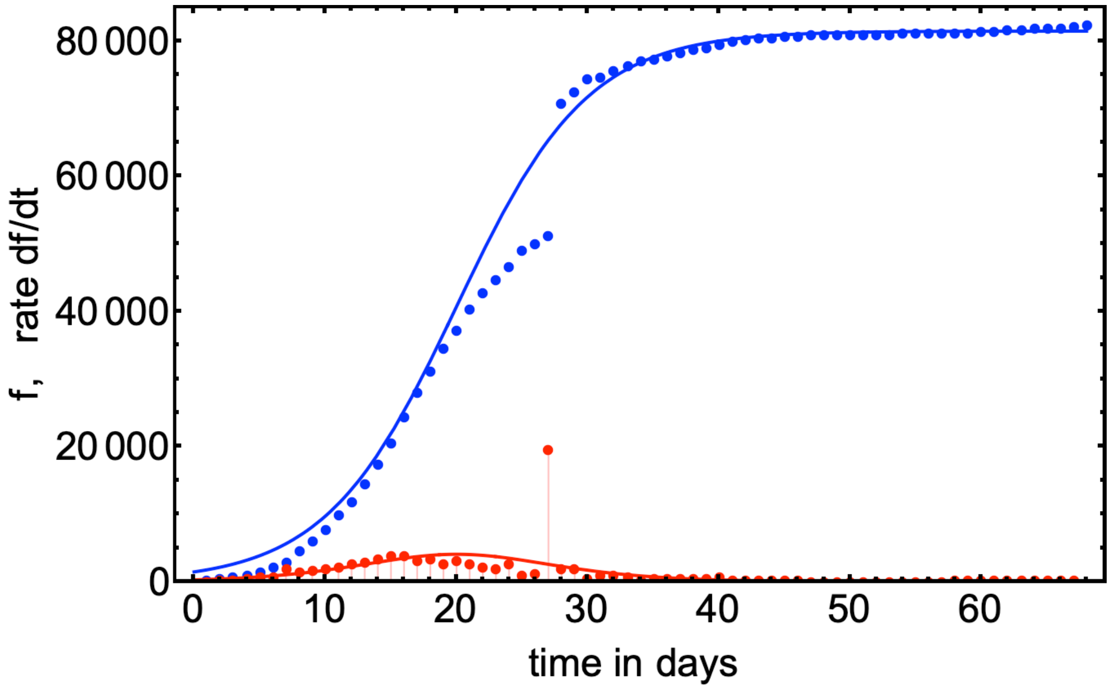

Figure 10: Comparison of WHO data for China with a logistic model.
Reported cumulative infections (blue dots) and daily rate of infection (red dots).
Computed logistic model for cumulative infections (blue line) and daily rate of infection (red line).
Logistic model parameters: $c_{max} = 81500$, $t_{peak} = 20$, $T = 5$ days.

3.3.2 Italy data

Figure 11 shows the WHO data (blue dots) for the cumulative infections and daily rate of infection (red dots) for Italy with superposed plots of a logistic function f for the cumulative infections and daily rate of infections df/dt. The logistic model parameters are $c_{max} = 130000$, $t_{peak} = 35.5$, $T = 5.8$ days. The logistic model shows a very good fit with an implied peak in the infection rate at 35.5 days.

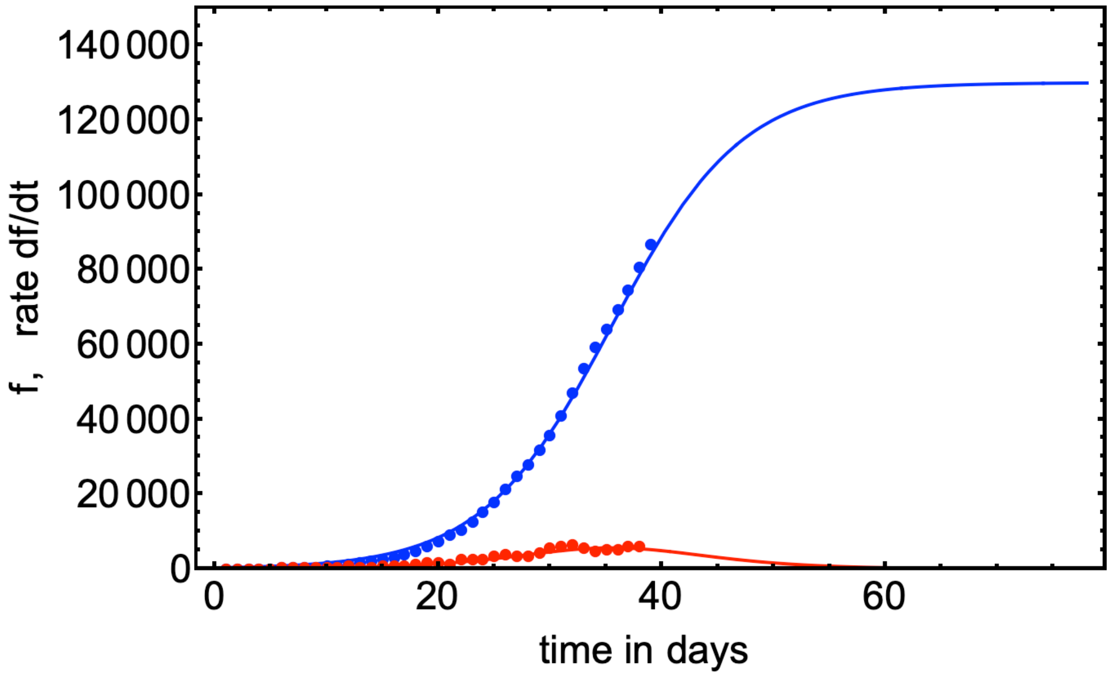

Figure 11: Comparison of WHO data for Italy with a logistic model.
Reported cumulative infections (blue dots) and daily rate of infection (red dots).
Computed logistic model for cumulative infections (blue line) and daily rate of infection (red line).
Logistic model parameters: $c_{max} = 130000$, $t_{peak} = 35.5$, $T = 5.8$ days.

3.3.3 Context: Global data

To set the previous results in context Figures 12 -13 show the WHO reported data for the global infections of COVID-19. The spike in the reported data was discussed in section 1. The data follows the multiple logistic picture described in section 3.2 with a first large outbreak in China followed later by European countries and final the rest of the world.

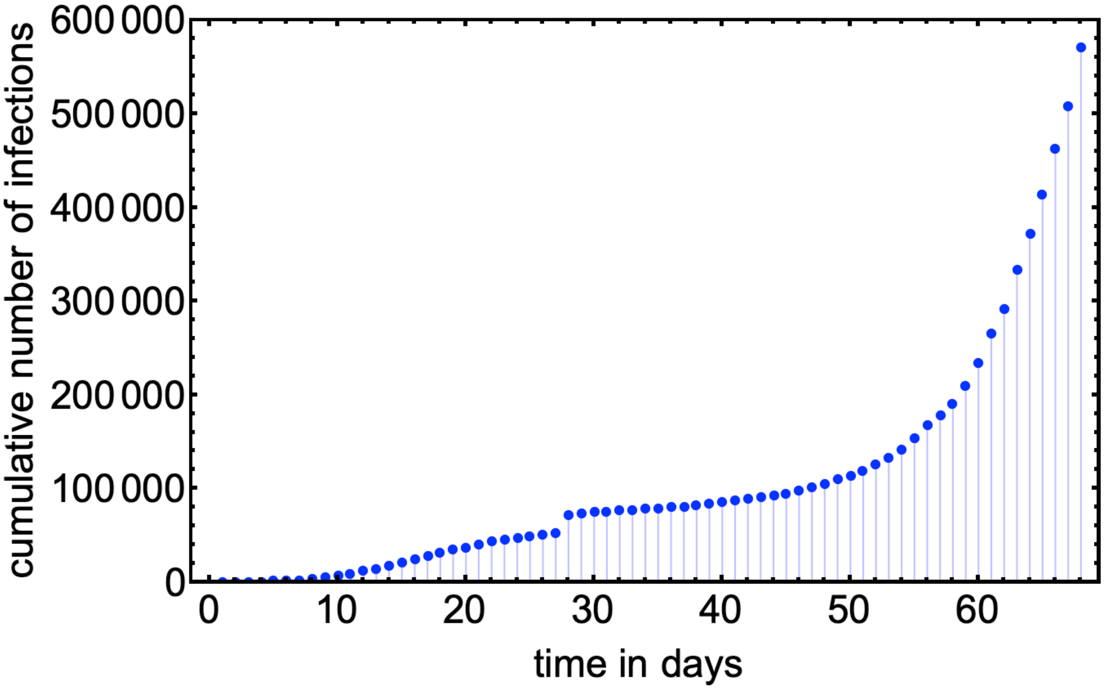

Figure 12: Global cumulative infections as reported by WHO. The jump at day 28 is due to a large discrepancy caused by a change in the counting methodology in China.

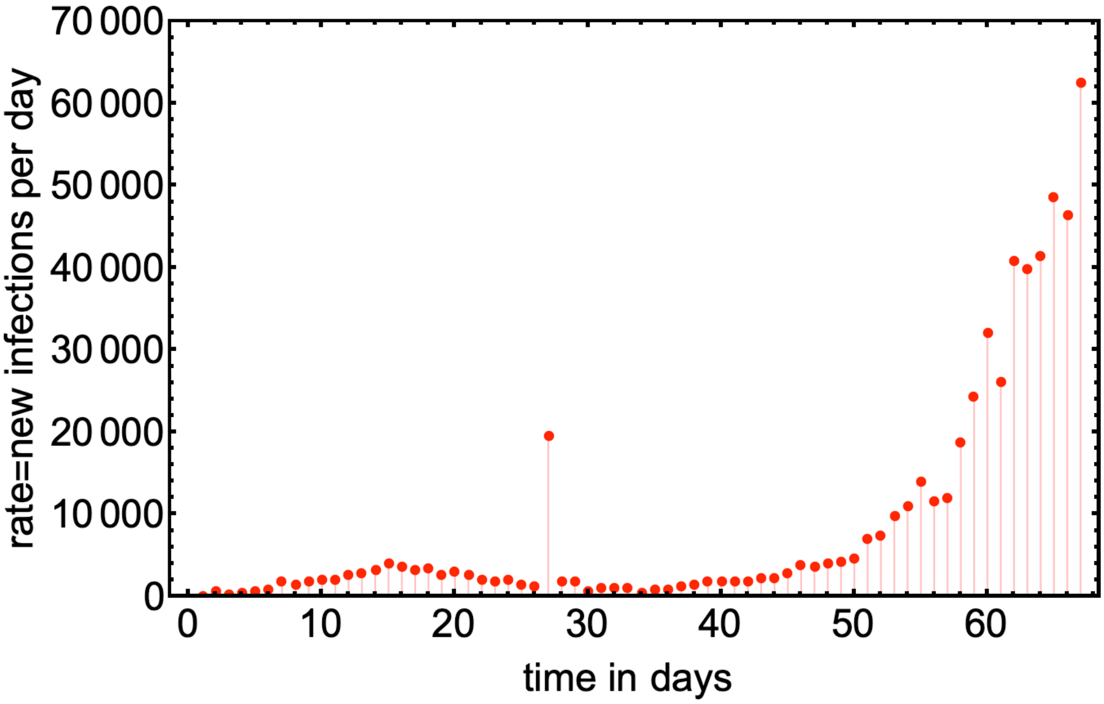

Figure 13: Global daily rate of infections as reported by WHO. The spike at day 28 is due to a large discrepancy caused by a change in the counting methodology in China.

4. Conclusions

The logistic function is a useful modelling tool with a small number of parameters that captures the statistical forms of a simple epidemic: in the present case the COVID-19 outbreak. By using superpositions of logistic functions it may be feasible to model global pandemics where multiple staggered outbreaks occur. The simple model gives a good fit to the data reported for China and Italy. The procedure outlined here is basically simple curve fitting but taking note of the physical origin of the parameters and noting the exponential phase of simple epidemics. This must be regarded as only an elementary step in understanding the spread of infections and their prediction. We refer to the extensive epidemiological literature elsewhere to see the full scope and complexity of this important field. The code for implementing the results presented here will be made available separately.

Note: the author is not an epidemiologist but has a background in theoretical physics.

References

1. World Health Organisation Situation Reports
https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

2. Moore, G., " Moore's law at 40", in *Understanding Moore's Law: Four Decades of Innovation*, Brock, D., (ed.), (Chemical Heritage Foundation) Chapter 7, 67–84, (2006).

3. Ritter, F. E., and Schooler, L. J., “The learning curve”, *International Encyclopedia of the Social and Behavioral Sciences*: Amsterdam: Pergamon 8602-8605, (2002).

4. Verhulst, P-F., "Recherches mathématiques sur la loi d'accroissement de la population", Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles,**18**, (1845).

5. Blakemore, J.S., *Semiconductor Statistics: International Series of Monographs on Semiconductors,Pergammon Press (2002).*

6. https://mathworld.wolfram.com/LogisticEquation.html

7. Barker, J.R., and Martinez , A.,“Image charge models for accurate construction of the electrostatic self-energy of 3D layered nanostructure devices”, Journal of Physics : Condensed Matter, **30** (13), 134002 (2018)

Appendix: Extracts from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

WHO Report No.	Date	Global Cumulative Infections	China Cumulative Infections	China Deaths	Italy Cumulative Infections	Italy Death	Italy infection rate
2			309				
3	23/01/20	581	571				
4		846	830				
5		1320	1297				
6		2014	1985				
7		2798	2741				
8		4598	4537	106			
9		6065	5597	132			
10		7818	7736	170			
11		9826	9720	213			
12	01/02/20	11953	11821	259			
13		14557	14411	304			
14		17391	17238	361			
15		20630	20471	425			
16		24554	24363	491			
17		28276	28060	564			
18		31481	31211	637			
19		34886	34598	723			
20		37558	37251	812			
21		40554	40235	909			
22		43103	42708	1017			
23		45171	44730	1114			
24		46997	46550	1368			
25		49053	48548	1381			
26		50580	50054	1524			
27		51857	51174	1666			
28		71429	70635	1772			
29		73332	72528	1870			
30	19/02/20	75204	74280	2006	3	0	
31		75748	74675	2121	3	0	0
32		76769	75569	2239	3	0	0
33		77794	76392	2348	9	0	6
34		78811	77042	2445	76	2	67
35		79331	77262	2595	124	2	48
36		80239	77780	2666	229	6	105
37		81109	78191	2718	322	11	93
38		82294	78630	2747	400	12	78
39		83652	78961	2791	650	17	250
40		85403	79394	2838	888	21	238
41	01/03/02	87137	79968	2873	1128	29	240
42		88948	80174	2915	1689	35	561
43		90869	80304	2946	2036	52	347
44		93091	80422	2984	2502	80	466
45		95324	80565	3015	3089	107	587
46		98192	80711	3045	3858	148	769
47		101927	80813	3073	4636	197	778
48		105586	80859	3100	5883	234	1247
49		109577	80904	3123	7375	366	1492
50		113702	80924	3140	9172	463	1797
51		118319	80955	3262	10149	631	977
52	12/03/20	125260	80981	3173	12462	827	2313
53		132758	80991	3180	15113	1016	2651
54		142534	81021	3194	17660	1268	2547
55		153517	81048	3204	21157	1441	3497
56		167515	81077	3218	24747	1809	3590
57		179111	81116	3231	27980	2503	3233
58		191127	81116	3231	31506	2503	3526
59		209839	81174	3242	35713	2978	4207
60		234073	81300	3253	41035	3407	5322
61		266073	81416	3261	47021	4032	5986
62		292142	81498	3267	53578	4827	6557
63		332930	81601	3276	59138	5476	5560
64	24/03/20	372757	81747	3283	63927	6077	4789
65		414179	81848	3287	69176	6820	5249
66		462684	81961	3293	74386	7505	5210
67		509164	82078	3298	80539	81675	6153
68	28/03/20	571678	82230	3301	86498	9136	5959