

Modification of the Logistic Model for Epidemics and Application to the COVID-19 Epidemic of 2020 in the United States

David H. Roberts¹

*Department of Physics MS-057
Brandeis University
Waltham, MA 02454-0911 USA
roberts@brandeis.edu*

Version of 2020 May 24

ABSTRACT

We present a modification of the logistic model of epidemics that takes into account the possibility that an epidemic can develop from multiple physically-distinct hot spots with a range of starting times. This produces an improved understanding of the time evolution of the COVID-19 epidemic taking place in the United States in the spring of 2020.

1. Introduction

In the spring of 2020 a pandemic of infection of a novel coronavirus (SARS-CoV-2) is overspreading the world. The United States in particular has been very hard hit by this disease (see, for example, the coverage in New York Times (2020)). Epidemiologists have struggled to describe adequately this pandemic and to predict its future course. This is a complex undertaking, involving the mathematics of epidemiology and a huge amount of uncertain data that serve as input to the models (see, for example, Fivethirtyeight (2020)). In this paper I examine the ability of a basic epidemiological model, the *logistic model*, to describe the course of the epidemic in the US. After showing that it is inadequate to the task, I present a straightforward improvement of the model, the *distributed logistic model*, that proves to work much better.

¹ORCID 0000-0001-5226-4540.

– 2 –

In the sections below I describe the logistic model and apply it to the US epidemic (§2), derive the distributed logistic model and show that it fits the data quite well (§3), and discuss the results (§4).

2. The Logistic Model

2.1. Derivation

A simple model for the evolution of an epidemic is based on the logistic differential equation. This describes an epidemic that begins with a small number f_0 of infected individuals at a single time and place, and subsequently spreads through a population. The motivation for this model is as follows. If the population of infected individuals as a function of time is $f(t)$, simple exponential growth with growth rate r is determined by

$$\frac{df}{dt} = rf,$$

with the growing exponential solution

$$f(t) = f_0 e^{rt},$$

where $f_0 = f(0)$. This is what happens with an infinite pool of subjects. However, for a finite pool of subjects, as the population of infected individuals grows the number of subjects available to be infected gets smaller. This is taken into account by modifying the exponential differential equation to become

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

where K is the total available pool of individuals. The solution of this *logistic equation*² is

$$f(t) = \frac{K f_0 e^{rt}}{K + f_0 (e^{rt} - 1)}, \quad (2)$$

which satisfies the required limits $f(0) = f_0$ and $f(\infty) = K$. The time course described by Equation 2 is the familiar “S curve” used to describe bacterial growth and other phenomena (see Figure 1).

An analysis of the total number of cases as a function of time $f(t)$ is just one way to compare model and data. Instead we can examine the number of new cases per day

²This is a version of the Bernoulli differential equation $f' = f(1 - f)$.

– 3 –

as a function of time; this is the time derivative of $f(t)$.³ This is most easily found by substituting Equation 2 into Equation 1,

$$\frac{df}{dt} = f_0 K r e^{rt} \left[\frac{K - f_0}{((e^{rt} - 1)f_0 + K)^2} \right]. \quad (3)$$

In either case, for the logistic model the parameters to be adjusted are f_0 , K , and r .

2.2. Fit to the COVID-19 Epidemic

When applied to the current COVID-19 epidemic in the United States the logistic model does only a fair job of accounting for the actual history of the total number of infected individuals as a function of time in the first half of 2020. In Figure 1 we show the data (Worldometers (2020)) for the US as of May 22, 2020,⁴ along with the best fit⁵ logistic function. In Figure 2 we show the differential counts of cases, that is, the number of new cases per day as a function of time. The best fitting parameters in each case are shown in Table 1. It is fair to say that the fits to the data expressed either way are rather disappointing. In particular, the logistic model is unable to account for the long linear portion of the curve of total cases, corresponding to a long period of approximate constancy in the daily counts.

³Fitting the daily numbers is statistically the preferred procedure as the data points are independent, unlike those for daily totals.

⁴Day 1 is March 23, 2020.

⁵The fits were made using the non-linear curve fitting capability of *Kaleidagraph*TM. See the Appendix for details.

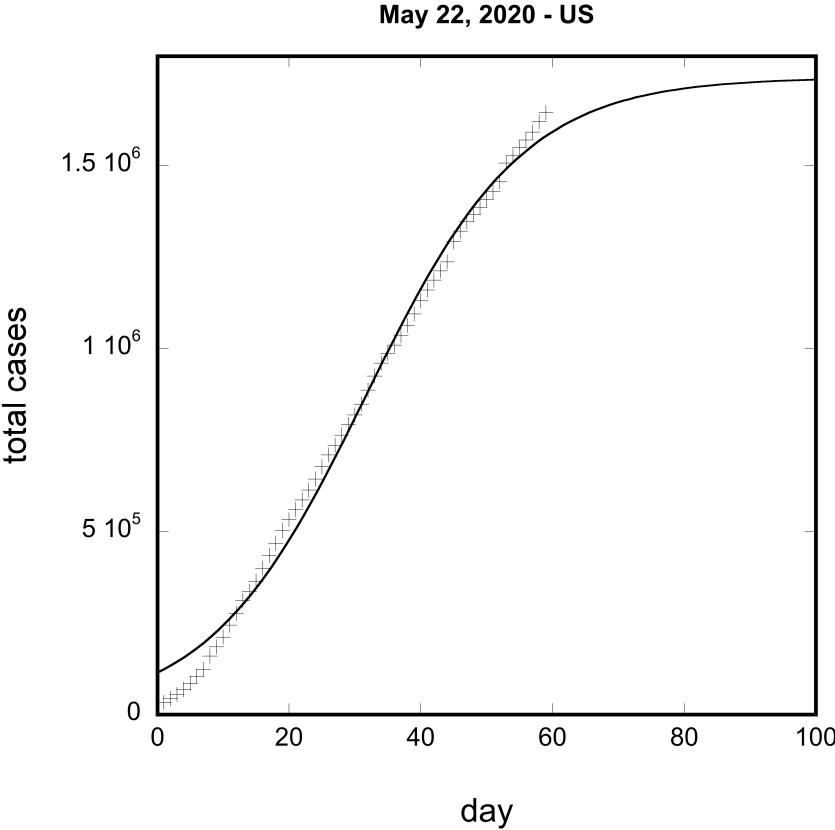


Fig. 1.— Fit of the logistic model to total US cases of COVID-19. The crosses are the data points and the curve is the fit to Equation 2; the best fit parameters are given in Table 1.

- 5 -

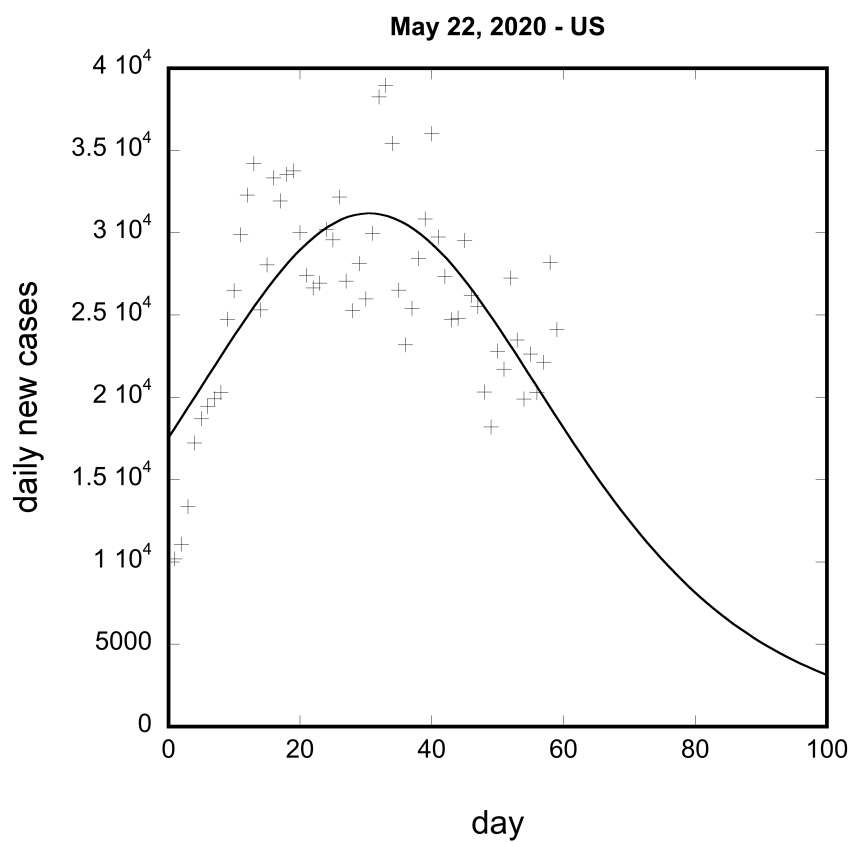


Fig. 2.— Fit of the logistic model to daily new US cases of COVID-19. The crosses are the data points and the curve is the fit to Equation 3; the best fit parameters are given in Table 1.

3. The Distributed Logistic Model

3.1. Derivation

In reality the 2020 COVID-19 epidemic in the US is a series of epidemics centered at various places distributed across the country and beginning over a range of times. We can crudely account for this by averaging a series of logistic distributions over a range of starting times t_i spanning a total time T ; the appropriate function to be averaged is

$$f(t) = \frac{K f_0 e^{r(t-t_i)}}{K + f_0 (e^{r(t-t_i)} - 1)}. \quad (4)$$

To account for a total range of time T over which the various hot spots of the epidemic are distributed we integrate Equation 4 over $0 \leq t_i \leq T$ and divide by T . The result is

$$g(t) = \frac{K}{rT} \ln \left[\frac{(e^{rt} - 1)f_0 + K}{(e^{r(t-T)} - 1)f_0 + K} \right] \quad (5)$$

Note that

$$g(0) = \frac{K}{rT} \ln \left[\frac{K}{(e^{-rT} - 1)f_0 + K} \right],$$

not f_0 , and that

$$g(\infty) = K.$$

From Equation 5 we can derive the differential distribution

$$\frac{dg}{dt} = \frac{f_0 K (f_0 - K)}{T} \frac{(e^{rT} - 1)}{((f_0 - K)e^{r(-t+T)} - f_0)(K + f_0(e^{rt} - 1))}. \quad (6)$$

For the distributed logistic model the parameters to be adjusted are f_0 , K , r , and T .

3.2. Fit to the COVID-19 Epidemic

In Figures 3 & 4 we show the best fits of the distributed logistic model to the cumulative and daily numbers. It seems fair to say that the results are significant improvements over what we get for the basic logistic model.

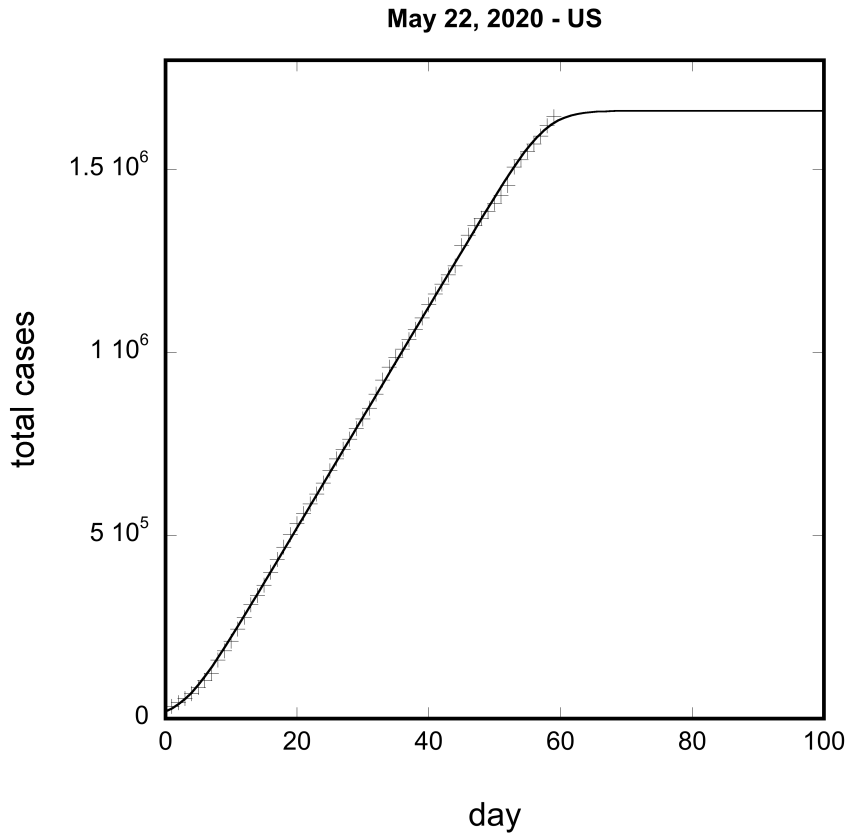


Fig. 3.— Fit of the distributed logistic model to total US cases of COVID-19. The crosses are the data points and the curve is the fit to Equation 5; the best fit parameters are given in Table 1.

- 8 -

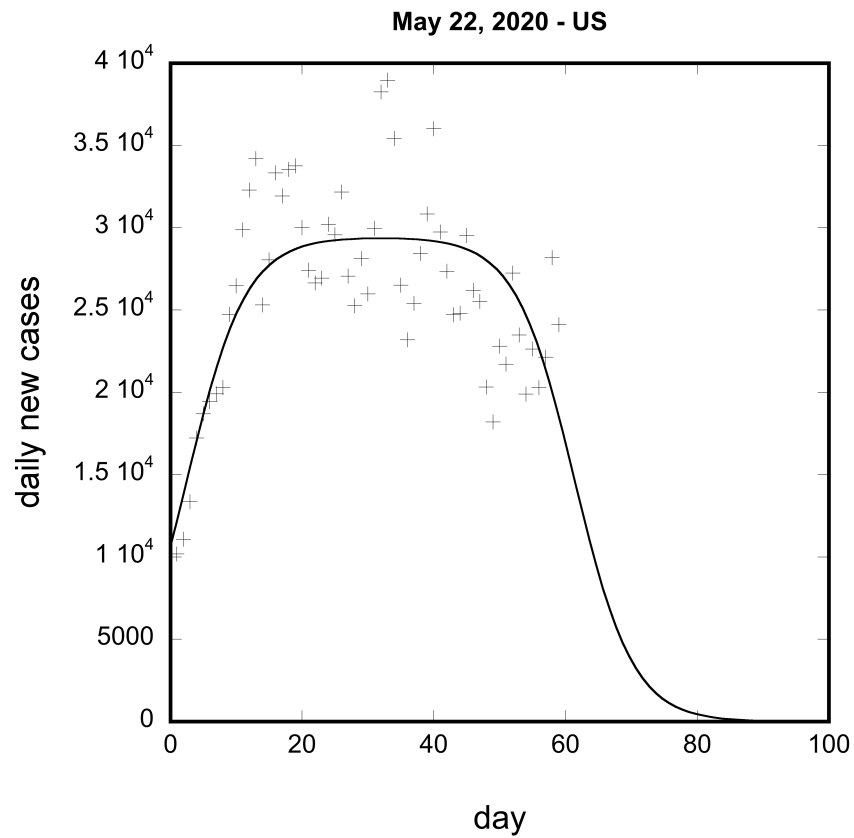


Fig. 4.— Fit of the distributed logistic model to daily new US cases of COVID-19. The crosses are the data points and the curve is the fit to Equation 6; the best fit parameters are given in Table 1.

Table 1. Parameters of the Fits

Parameter	$N(t)$	$dN(t)/dt$	$N(t)$	$dN(t)/dt$
Model	Logistic	Logistic	Distributed Logistic	Distributed Logistic
Equation	(2)	(3)	(5)	(6)
f_0	1.1×10^5	4.0×10^5	4.0×10^5	6.1×10^5
K	1.7×10^6	2.4×10^6	1.7×10^6	1.7×10^6
r (days ⁻¹)	0.084	0.052	0.42	0.23
T (days)	55	59

4. Discussion

We have demonstrated that a useful way to improve the logistic description of the evolution of the 2020 COVID-19 epidemic in the US⁶ is to average logistic functions over a range of starting times (T). The concept of distributed start times in an epidemic makes sense for all but the most compact geographies. If you think of the approximately 3000 counties in the US starting up at various times over a couple of months, you get the picture of a set of 3000 S-shaped curves of cases each with different heights and start dates. Averaging those curves produces one aggregated curve (Equation 5).

This approach could be modified by introducing a suitable weighting function for the starting times. While such calculations can be done for some reasonable weighting functions,⁷ the number of additional parameters required would seem to make this modification not useful.

The new parameter of the distributed logistic model is the time span T , which turns out to be about 57 days. Note that this corresponds closely to the full width at half maximum of the differential curve in Figure 4. It is significant that the integral and differential distributions lead to the same values. These numbers are also consistent with the history of the spread of the epidemic from hot spot to hot spot across the US (see, for example, the coverage in the New York Times (2020)).

The rate of infection r translates to times between infections due to a single individual $\tau = 1/r$ (estimated from fitting using the distributed logistical model) between about 2 and 4 days. If the pre-symptomatic period of initial infection is Δt , then the number of individuals infected by a single pre-symptomatic infected individual is about $\Delta t/\tau = r\Delta t$. If $\Delta t \simeq 10$ days, then *on the average* each individual infects about three others before discovering that he/she is infected. This product $r\Delta t$ is approximately the parameter R_0 of a standard SIR epidemiological model.

Is this model capable of predicting the future of this epidemic? We have explored this possibility, and conclude that the real world data do not constrain the parameters of the model sufficiently well for this to be reliable for the 2020 COVID-19 epidemic in the United States. As we have continued to include more data the parameters of the best fits have evolved; we conclude that at the current time hot spots are continuing to develop across

⁶We have not attempted to apply the model to any individual state, to any other country, or to the world in aggregate.

⁷In particular, both Gaussian and triangular distributions.

the country, so that the time T is increasing.⁸ This means that the model is suitable for description at a given moment, but not for prediction until the best fit parameters stop evolving, that is, until new hot spots stop appearing across the country.

Finally, note that this model is not a replacement for sophisticated epidemiological modeling, but is a simple way to see where the basic logistic model goes wrong, and how it can be improved.

5. Acknowledgements

We thank Brian Boyle, Mary Roberts, and Bob Sauer for their helpful comments and suggestions.

6. References

Fivethirtyeight (2020), <https://projects.fivethirtyeight.com/covid-forecasts/>.

New York Times (2020), multiple news stories, New York, New York.

Worldometers (2020), <https://www.worldometers.info/coronavirus/country/us/>.

⁸Fitting the first 30 days of data yields $T \simeq 30$ days.

7. Appendix

The translation of Equations 2, 3, 5, & 6 into Kaleidagraph™ code is as follows. The mapping of the parameters is

$$t \rightarrow M0,$$

$$f_0 \rightarrow M1,$$

$$K \rightarrow M2,$$

$$r \rightarrow M3$$

$$T \rightarrow M4.$$

The translated equations are

$$f(t) \rightarrow \frac{e^{M0M3}M1M2}{(-1 + e^{M0M3})M1 + M2}$$

$$\frac{df}{dt} \rightarrow \frac{e^{M0M3}M1M2(-M1 + M2)M3}{((-1 + e^{M0M3})M1 + M2)^2},$$

$$g(t) \rightarrow \frac{M2\text{Ln} \left[\frac{M1 - e^{M0M3}M1 - M2}{M1 - e^{M3(M0 - M4)}M1 - M2} \right]}{M3M4},$$

and

$$\frac{dg}{dt} \rightarrow \frac{(-1 + e^{M3M4})M1(M1 - M2)M2}{(-M1 + e^{M3(-M0 + M4)}(M1 - M2))((-1 + e^{M3M0})M1 + M2)M4}.$$