 doi:10.20944/preprints202008.0479.v1

Epidemic Evolution: Multiple Analytical Solutions for the SIR Model

Ian Lerche*

Institut für Geowissenschaften, Naturwissenschaftliche Fakultät III, Martin-Luther-Universität, Germany

**email: lercheian@yahoo.com*

Abstract

While there are many models of epidemic evolution perhaps the basis for such models finds itself in the lumped behavior expressed through the so-called SIR model (Susceptible, Infectious, Recovered) from which spring many related models. This paper discusses multiple analytic solutions to that equation including those that are available in closed analytic form and those for which at least one final integral has to be done numerically, so-called quasi-analytic solutions. The solutions are intrinsically time-dependent of course. The hope is that such an investigation will lead to a better understanding of when and how models can be of use in studying the dynamical evolution of diseases including, perhaps, the great influenza pandemic of 1918 together with later pandemics and epidemics not excluding the Covid-19 pandemic of the present day.

Keywords: SIR model, epidemic, multiple analytical solutions.

Introduction

It is now almost 100 years since a concerted effort was undertaken to describe quantitatively the temporal evolution of epidemic and pandemic diseases (Kermack and McKendrick, 1927) perhaps triggered by the devastating pandemic of the so-called Great Influenza which ravaged the world for over three year starting around 1918 and estimates put the world death toll at around 50-100 million people (Barry, 2004), although there is considerable uncertainty on even these estimates. Since then there have been numerous quantitative models suggested to describe the evolution of infectious diseases. A good summary is provided by Hethcode (2000) covering a wide variety of behaviors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2020 doi:10.20944/preprints202008.0479.v1

Despite the progress made with such models, including their strengths and weaknesses, one is nevertheless in the situation where most model behaviors have been undertaken with numerical procedures so that one does not see so transparently the general structure of solutions and initial conditions dependency without an enormous numerical effort. The main difficulty is not so much in the mathematical equations *per se* but in the use of "blocks" to describe the epidemiology without due concern being given to the underlying causes of the block parameters. Perhaps such a description is inevitable but it leaves one with less than complete satisfaction with such models.

Accordingly here we revert to the basic model of 1927 (Kermack and McKendrick, 1927) to investigate the solutions behavior for analytic models. i.e. models that can be completed in closed analytic form so that there is no influence of any numerical algorithm. The major point to be addressed is to see to what extent solutions are influenced by both the combined choices of parameters in the blocks and the initial conditions. The non-linearity of the basic equations is the main cause of the interaction of parameter choices and initial conditions and leads to constraints that would not otherwise obtain. Such an investigation points the way to a more complete description of model behaviors and so indicates the potential for obtaining quantitative results more closely in accord with observations.

Technical Development

Consider the basic equation (Kermack and McKendrick, 1927)

$$I+S+R=1 \quad (1)$$

where I, R, S are the relative fractions of infected, recovered and susceptible people involved at time t. Note that each of I, S and R is a fraction and each must be not only positive but also less than unity. In addition note that this basic model does not break out that fraction of the population that dies nor does it allow for a change in the base population due to immigration, emigration, non-epidemic death, immunity or birth. Nevertheless the behavior provides a useful block model from which one can springboard to more erudite models should one choose although, as will soon be shown, there are enough uncertainties within the basic model to raise concerns about the worth increase of more sophisticated models.

Following Kroeger and Schlickeiser (2020) let a(t) and μ(t) denote the infection and recovery rates, with both positive, then the dynamical block model is given through

$$dI/dt = a(t)SI - \mu(t)I \quad (2)$$

$$dS/dt = -a(t)\,SI \quad (3)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2020 doi:10.20944/preprints202008.0479.v1

together with the constraint (1) .The reason for calling this model a block model is because the whole of the epidemiology is contained in the lumped factors a(t) and μ(t) without being more specific concerning, for example, asymptomatic effects, gender, race, age or recurrence rates. Nor indeed is any spatial variation considered so that all fractions of a country are handled as though they were all infected equally at a given time.

The nonlinearity of equations (2) and (3) makes direct attempts to obtain solutions less than easy. In addition once one has solutions then one must address the problems of initial conditions for it is within the framework of chosen initial conditions that one is interested in solution behaviors. There is also the question of the values to use for the two lumped factors a(t) and μ(t) so that solutions to the model equations can mirror as closely as possible the actual behavior of an epidemic or pandemic disease. Alternatives exist of course.

One could measure directly S, I and R at each instance of time and then invert equations (2) and (3) to write

$$a(t) = dS/dt /(SI) \quad (4)$$

and

$$\mu(t) = (dS/dt - dI/dt)/I \quad (5)$$

thereby determining the values as direct functions of time. While such may be more appealing to some the concern is that the data one needs to compute a(t) and μ(t) should be complete, accurate and precise. Such precision is not likely as evidenced by information for the great influenza pandemic of 1918 (Barry, 2004) where even to this day it is not known how many people died except for rough estimates of between 50 million and 100 million so that one can be duly skeptical of data made available.

Accordingly interest here will center on determining analytic solutions to the relevant equations for particular behaviors of a(t) and μ(t). To this end write the ratio of equations (2) and (3) to obtain

$$dI/dS = -1 + k(S)/S \quad (6)$$

where $k(S) = (\mu(t)/a(t))$ and one uses S as a basic variable rather than time t. Note that only the ratio k(S) enters equation(6) so enabling considerable progress to be made with determining solutions. Perhaps the most obvious situation is when one considers k(S) to be constant, K say, although individually a(t) and μ(t) may vary but in fixed ratio.

The general solution to equation (6) is then

$$I(S) = J - S + K\ln S \quad (7)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2020 doi:10.20944/preprints202008.0479.v1

where J is a constant. The initial condition that there be no infections (I = 0) at time t =0 (correspondingly there can be no recoveries at that time if no-one has been infected) and so one has S= 1 at time t = 0 with then J = 1 so the complete solution in this case is

$$I(S) = 1-S+K \ln S \tag{8}$$

Now as time progresses one has S decreasing but the limit is when I = 0 because smaller values of S would then give negative values for I which are intrinsically forbidden. For specified K the limiting value of S, S*, is then given through

$$\ln S^* = -(1-S^*)/K \tag{9}$$

indicating a residual population of susceptible people. Figure 1 shows a plot of equation (9).

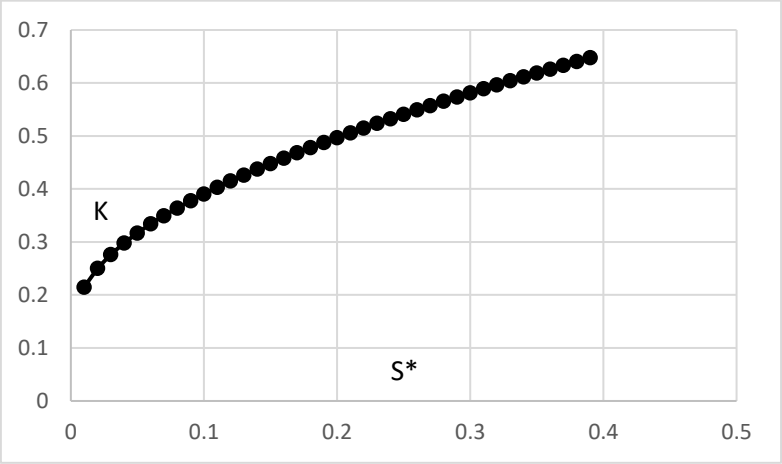

Figure 1. Plot of the critical value S* versus K from equation (9)

But this solution for k(S) =K =constant is not analytic in all variables because the determination of S as a function of t is found through

$$dS \left(S(1-S+K \ln S)\right)^{-1} = -a(t)dt \tag{10a}$$

which must be done numerically. Thus the designation quasi-analytical.

The choice of k(S)=K = constant was completely arbitrary and there seems to be no fundamental reason that a disease should choose to honor this requirement (there seems to be no obvious reason a disease is forbidden to have k(S) constant either!). It is, therefore, germane to consider other situations for they more be more realistic in their approach to understanding how well the basic model can pattern itself to observed disease behaviors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2020 doi:10.20944/preprints202008.0479.v1

Generally then one has

$$I(S) = 1 - S - \int k(S')/S' dS' \quad (10b)$$

where the upper (lower) integral limits are 1 (S) respectively and I(S = 1) = 0.One also has

$$dS/dt = -a(t)\, S\, (1 - S - \int k(S')/S' dS') \quad (11)$$

with the recovery rate given through R =1-I-S.

Simple solutions for variable k(S) are readily available in totally analytic form. For instance if k(S) = LS with L constant then I(S) = 1-S-L(1-S) or if k(S) = M S^2 then I (S) = 1-S –M(1-S^2)/2 and so on through k(S) = $P_n S^n$ when I(S)= 1-S –P_n(1-S^n)/n. Indeed linear addition of such powers can also be undertaken to yield analytic results for I(S). However note that whatever one chooses one always has to have 0<I(S)<1. It is also possible to discuss situations in which one takes k(S) to be represented by negative powers of S such as $Q_n S^{-n}$ and also of course linear combinations of negative powers.

However, in order to obtain a complete closed form analytical solution for all components some care needs to be exercised. In order to ensure an analytic closed form result for *both* I(S) and S(t) one sees immediately that one must restrict a polynomial behavior for k(S) to polynomials of order S^4 or less in the case of positive powers and equally so for negative powers . The reason is that in order to factor (1 -S - $\int$k(S')/S'dS') so that one can perform the integration of dS/dt in closed form one cannot generally factor polynomials higher than degree S^4 .

To illustrate behaviors different than for the case of constant k(S) consider the situation where k(S) is linear with k(S) =FS where F is a positive constant. Then one has

$$I(S) = (1-S)(1-F) \quad (12)$$

so that only for F <1 is there the chance of a solution. With equation (12) one then has

$$dS/dt = -a(t)\, S\, (1 - S)(1-F) \quad (13)$$

which has the general solution

$$\ln(S/(1-S)) = A-(1-F)T, \quad (14)$$

with dT =a(t)dt and A an integration constant.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2020 doi:10.20944/preprints202008.0479.v1

One can rewrite equation (14) in the more palatable form

$$S = \exp (A-(1-F)T)/(1+\exp(A-(1-F)T)) \quad (15)$$

However note that on T= 0 one has $S_0 = \exp(A)/(1+\exp(A))$ so that S does not reach unity for any finite A. Thus one does not start with a population of 100% uninfected people but rather with a smaller fraction S_0 . In short the initial conditions one wishes to impose are not acceptable; they have to be modified.

Thus one sees in both the case of constant k as well as that of k linear in S the solutions to the evolution equations limit what one can say and do and are not immune to the desired initial conditions. In short what is shown is that both the general structure of the equations as well as the initial conditions are beholden to the choice one makes for k(S). In addition analytic solution of the equations depends on whether one can truly provide a complete analytic solution or whether one is reduced to performing integrals numerically. In the latter case one has a quasi-analytic solution but also with restrictions.

Should one wish to proceed further and solve the evolution equations for higher powers of k dependent on S then one again has restrictive behaviors as is also the case when one considers linear combinations in the sense of $k(S) =AS +BS^2 +CS^3+DS^4$ with A,B,C and D constants but such that k(S) is positive.

An example of the constraint is illustrated for the case where one chooses $k(S) = AS+BS^2$. Because k(S) must be positive in $0<S<1$ it follows that $A>0$ and $A>-B$. Thus B can be negative as long as A>mod(B).

Under this choice for k(S) one obtains

$$(1 -S - \int k(S')/S'dS') = (1-S)(1-A-B/2 -BS/2) \quad (16)$$

Then from equation (11) it follows that

$$dS/(S*(1-S)*(1-A –B/2 –BS/2)) =-a(t)dt \quad (17)$$

with solution in implicit form

$$S = \exp(-L-T)/(g(S)+\exp(-L-T)) \quad (18)$$

where L is a constant of integration, T is a time variable obtained from $a(t)(1-A-B/2)dt =dT$ once one has specified $a(t)>0$, and where

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2020 doi:10.20944/preprints202008.0479.v1

$$g(S) = ((1-A-B/2)/ (1-A-B/2-BS/2))^{p} *(1-S)^{-p} \quad (19)$$

with p =(B/2)*(1-A-B)/ (1-A-B/2).

Note that if B = 0 then p = 0 and equation (18) reduces to the solution given through equation (15). But now consider that B is non-zero and then consider the behavior as S tends to unity-supposedly on T = 0. If B <0 and because A>mod(B) under all conditions then near S = 1 one has $g(S\rightarrow 1) = (1-S)^{-p}$ which tends to zero because p <0 so that S→1 as required. However if B>0 then as S→1 one obtains the limiting behavior from $g(S\rightarrow 1) = (1-S)^{-p}$ which tends to infinity because p>0 so that S tends to 0 instead of the supposed unity. Thus the initial conditions cannot be satisfied and so only the solution with B<0 is appropriate.

The point about this example is that the freedom to have ranges of parameters in a multi-parameter analytical solution is limited the moment one imposes the initial conditions. In the simple case here only after one has obtained the complete general solution can one determine that particular ranges of a parameter are not allowed.

While many such calculations are relatively simple to perform (and so are not included here) the points made also apply – often solutions are available only under restricted ranges of parameters, solutions do not always satisfy the presupposed initial conditions for the start of a pandemic, and so on. In addition there is no known fundamental reason to suppose a disease should obey the requirements discussed here for analytic solutions to the evolution equations. After all what does a disease know about higher mathematics?

A further point is relevant. One can often determine the number of infections from a pandemic disease as a function of time with somewhat better resolution than one can determine either the recovery rate or the susceptibility rate particularly when one is often faced with asymptomatic situations in people. Suppose for instance there were no recoveries. Then μ(t) = 0 and the general solution to the evolution equations is simply

$$I(t) = \exp(A+T)/(1+\exp(A+T)) \quad (20)$$

with A a constant of integration and dT =a(t)dt with, of course, T positive.

Then note that one cannot have I =0 on t = 0 instead one has I(0) =exp(A)/(1+exp(A)) which is intrinsically less than unity but not zero. As T becomes large without limit I(t) tends to unity so that S tends to zero and the disease infects (kills?) everyone. So one starts with a percentage of the population infected and life gets worse with increasing time when there are no recoveries-something that seems obvious in hindsight of course. The point being made is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2020 doi:10.20944/preprints202008.0479.v1

that the initial conditions depend on how parameters are chosen and the consequent solution of the equations thereafter. One cannot choose parameters and initial values independently, they influence each other.

It is also germane to note that a pandemic must start at some time, say t=0, so that one cannot have infected cases before that start time. Thus if one were to try to represent the infection rate with any model distribution that had a "tail" extending to values t<0 then that must at best be an approximation with, however, the possibility that the tail could have a major influence on the model results. Far better is to restrict one's attention to model distributions that exist only for t equal to and greater than zero so that there is no chance of such possibilities. For example one such model is the logarithmic distribution

$$P(t) = (tv\pi^{1/2})^{-1}\exp(-(\ln(t/r))^2/v^2) \qquad (21)$$

where t is time and r and v are constants. A plot of P(t) (in %) versus time (arbitrary units) for selected values of r and v is given in Figure 2

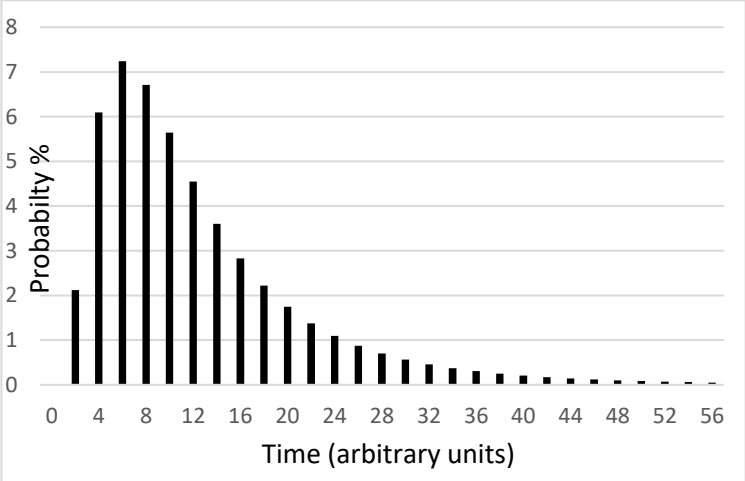

Figure 2. Plot of the log normal distribution P(t) versus time from equation (21).

One sees immediately that the plot restricts the cases of infection to positive values of time, as is required, and that the plot also shows a sharp rise in cases before a long slow drop as observed in reality (This shaping representation is far superior to assessments based on the assumption of a symmetric rise and fall of infected cases around the peak value of infections and is also much more in accord with observed patterns of rise and fall).

One can also investigate the cumulative probability with time. The logarithmic model is given in figure 3 and compares very favorably to the observed patterns of behavior for several European countries as given in Figure 2 of Lauer et al (2020). The similarity in the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2020 doi:10.20944/preprints202008.0479.v1

model and observed behaviors is striking and could be made even more so if one were to multiply the model behavior by the total number of observed cases or if one were to optimize the parameters v and r in equation (21) to agree with the observed behavior as closely as possible.

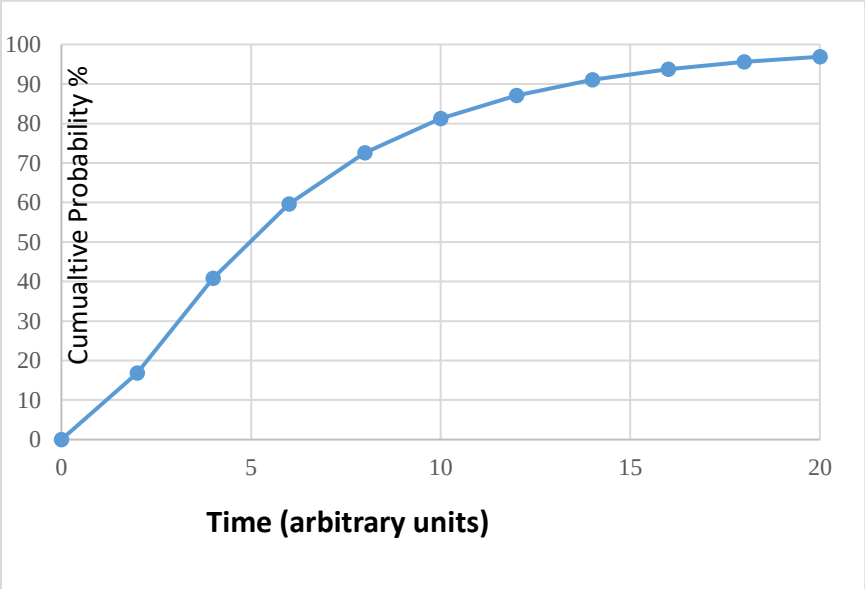

Figure 3. Model behavior for the cumulative probability (%) versus time based on the log normal distribution which compares very favorably to Figure 2 of Lauer et al. (2020) which, in turn, is based on a summary of observed symptomatic infections (fraction) versus days since infection for a variety of countries.

Conclusions

What has been shown here is that the behavior of analytical solutions to the SIR equations is highly dependent on the structure used for a single function entering the equations and the interlacing of parameters invoked is strongly coupled to the initial conditions required. There seems to be no obvious procedure available at the moment for determining the situations where one can effect an analytical solution from those where one cannot except for trial and error-a somewhat time consuming and not very transparent method.

There are alternative methods of defining epidemic or pandemic patterns such as inverse methods to determine functional behaviors from real data or logarithmic type distributions to quantify the evolution of a disease. However such methods are at best

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2020 doi:10.20944/preprints202008.0479.v1

empirical and one would like a more logical approach that covers all such eventualities. To date there is, apparently, no consensus on such a procedure. Until such time as one is available perhaps the best that can be undertaken is to see what individual calculations allow one to say. That theme has been the motivation behind the present work and will likely continue well into the future with the hope that such will eventually lead to a better understanding of when and how models can be of use in studying the dynamical evolution of diseases.

Appendix

While the SIR model of epidemic evolution is useful it has one very major drawback. As seen from the dynamical block model given through

$$dI/dt = a(t)SI - \mu(t)I \quad (A1)$$

$$dS/dt = -a(t)\,SI \quad (A2)$$

the evolution of infected people can proceed only if , at t = 0, there is a seed population of infected people because if I = 0 on t = 0 then it remains so forever after. Hence the need for a seed population of infected people. Indeed, there will also be no increase at time t= 0 in the fraction of the population infected if $S(0) < \mu(0/a(0)$ so that the fraction of susceptible people must also be large enough.

The major problem is that one must have a seed pool of infected people. One cannot argue that such a seed population came from some earlier time for one could then just reset the clock to that time and so again one would face the same dilemma. The seed population of infected people must be added to the SIR model in an ad hoc manner. One could, perhaps, argue that incorporating other effects such as gender, spatial variation, immunity, age distribution, and so on would mitigate the dilemma but then one would be dealing with a much more complex model with even more parameters and such a modified model would have its own weaknesses. While such a complexity is arguably unavoidable it is well beyond the description of the SIR model and any such development is best left to those more able than I. The point being made here is that extreme care has to be taken when discussing a model representation of epidemic evolution in order that one does not end up drawing conclusions not warranted by either the underlying limiting behavior incorporated in the model nor in attempts to apply a model with serious weaknesses to real life data.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

 doi:10.20944/preprints202008.0479.v1

References

Barry, J. M., 2004, *The Great Influenza*, Viking Press, New York, New York, 546 p.

Hethcode, H. W., 2000, The Mathematics of Infectious Diseases, SIAM Review, **42**, 599-653.

Kermack, W. O., and McKendrick, A. G., 1927, A contribution to the mathematical theory of epidemics, Proc. Roy. Soc. A , **115**, 700-721.

Kröger, M., and Schlickeiser, R., 2020, Analytic solution of the SIR-equation for the temporal evolution of epidemics. Part A: Time-independent reproduction factor. Preprint.

Lauer, S.A., Grantz, K.H., Bi O., and Jones, F.K., 2020, The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application, Annals of Internal Medicine **172,** 577-582