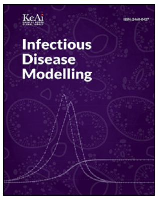
Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm



Estimating epidemic exponential growth rate and basic reproduction number

Junling Ma

Department of Mathematics and Statistics, University of Victoria, Victoria, BC, V8W 2Y2, Canada

ARTICLE INFO

Article history:
Received 17 December 2019
Accepted 27 December 2019
Available online 8 January 2020
Handling Editor: Dr. J Wu

Keywords:
Epidemic curve
Exponential growth rate
Maximum likelihood estimation
Phenomenological models

ABSTRACT

The initial exponential growth rate of an epidemic is an important measure of the severeness of the epidemic, and is also closely related to the basic reproduction number. Estimating the growth rate from the epidemic curve can be a challenge, because of its decays with time. For fast epidemics, the estimation is subject to over-fitting due to the limited number of data points available, which also limits our choice of models for the epidemic curve. We discuss the estimation of the growth rate using maximum likelihood method and simple models.

This is a series of lecture notes for a summer school in Shanxi University, China in 2019. The contents are based on Ma et al. (Ma, Dushoff, Bolker, & Earn, 2013). We will study the initial exponential growth rate of an epidemic in Section 1, the relationship between the exponential growth rate and the basic reproduction number in Section 2, an introduction to the least square estimation and its limitations in Section3, an introduction to the maximum likelihood estimation in Section 4, and the maximum likelihood estimation of the growth rate in Section 5.

1. Epidemic exponential growth rate

Epidemic curves are time series data of the number of cases per unit time. Common choices for the time unit include a day, a week, a month, etc. It is an important indication for the severeness of an epidemic as a function of time. For example, Fig. 1 shows the cumulative number of Ebola cases during the 2014–16 Ebola outbreak in western Africa. The cumulative cases during the initial growth phase form an approximately linear relationship with time in log-linear scale. Thus, in linear scale, the number of deaths increases exponentially with time. The mortality curve (the number of deaths per unit time) shows a similar pattern, as demonstrated by the daily influenza deaths in Philadelphia during the 1918 influenza pandemic shown in Fig. 2.

In fact, most epidemics grow approximately exponentially during the initial phase of an epidemic. This can be illustrated by the following examples.

Example 1. Consider the following Susceptible-Infectious-Recovered (SIR) model:

E-mail address: junlingm@uvic.ca.

Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2019.12.009

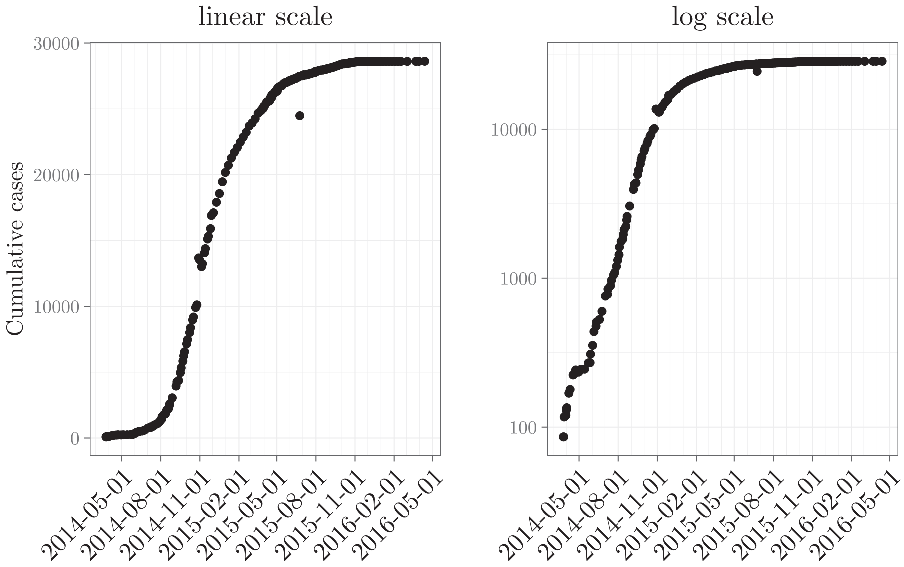


Fig. 1. Cumulative Ebola cases during the 2014–16 western African Ebola outbreak, plotted in linear scale (left) and log-linear scale (right). Source: Center for Disease Control Ebola case counts (Center for Disease Control, 2016).

$$\frac{dS}{dt} = -\beta SI \, , \tag{1a}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \, , \tag{1b}$$

$$\frac{dR}{dt} = \gamma I \tag{1c}$$

where S is the fraction of susceptible individuals, I is the fraction of infectious individuals, and R is the fraction of recovered individuals; β is the transmission rate per infectious individual, and γ is the recovery rate, i.e., the infectious period is exponentially distributed with a mean $1/\gamma$. Linearize about the disease-free equilibrium (DFE) $(1,0,0)$,

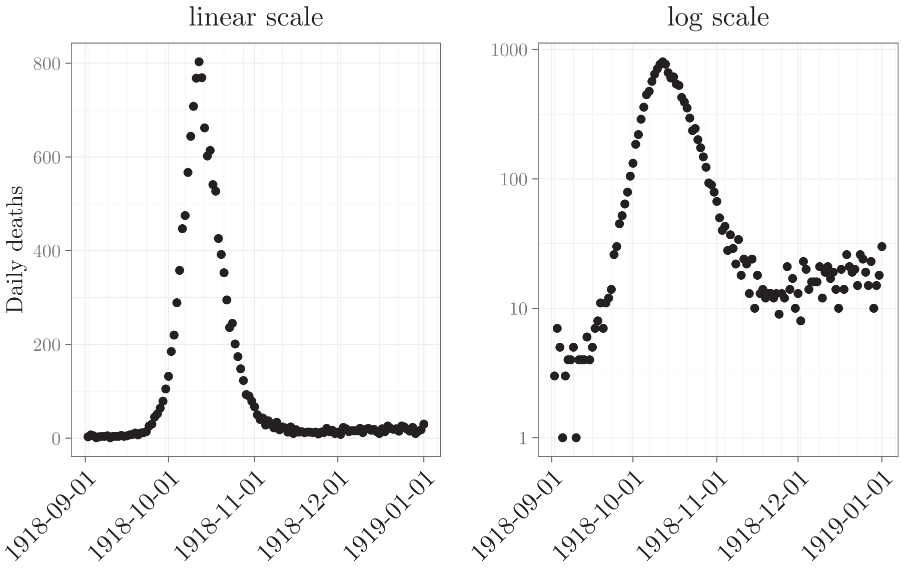


Fig. 2. Weekly influenza mortality during the 1918 pandemic in Philadelphia, plotted in linear scale (left) and log-linear scale (right).

$$\frac{dI}{dt} \approx (\beta - \gamma)I. \tag{2}$$

Thus, if $\beta - \gamma > 0$, then $I(t)$ grows exponentially about the DFE. In addition, initially, $S \approx 1$, thus, the incidence rate (number of new cases per unit time) $C = \beta SI$ also increases exponentially.

It is similar for an Susceptible-Exposed-Infectious-Recovered (SEIR) model, as illustrated by the following example.

Example 2. Lets consider an SEIR model:

$$\frac{dS}{dt} = -\beta SI, \tag{3a}$$

$$\frac{dE}{dt} = \beta SI - \sigma E \tag{3b}$$

$$\frac{dI}{dt} = \sigma E - \gamma I, \tag{3c}$$

$$\frac{dR}{dt} = \gamma I, \tag{3d}$$

where E is the fraction of latent individuals (infected but not infectious), σ the rate that latent individuals leaving the class, i.e; , the mean latent period is exponentially distributed with mean $1/\sigma$; S, I, R, β and γ are similarly defined as in Example 1. Again, $(1, 0, 0, 0)$ is a disease free equilibrium representing a completely susceptible population. Linearize about this equilibrium, the equations for E and I are decoupled, and become

$$\frac{dE}{dt} = -\sigma E + \beta I,$$

$$\frac{dI}{dt} = \sigma E - \gamma I.$$

Note that the Jacobian matrix

$$J = \begin{bmatrix} -\sigma & \beta \\ \sigma & -\gamma \end{bmatrix}$$

has two real eigenvalues, namely,

$$\lambda_1 = \frac{-(\sigma+\gamma) + \sqrt{(\sigma-\gamma)^2 + 4\sigma\beta}}{2}, \quad \lambda_2 = \frac{-(\sigma+\gamma) - \sqrt{(\sigma-\gamma)^2 + 4\sigma\beta}}{2}.$$

Thus, about the DFE, the solution of the model is asymptotically exponential with a rate λ_1. Similar to Example 1, the incidence rate also grows exponentially initially.

In general, suppose the infection states of an individual can be characterized by the following vector $(\vec{S}, \vec{I})$, where $\vec{S}$ represents multiple susceptible states, and $\vec{I}$ represents multiple infectious (or latent) states. We also use $\vec{S}$ and $\vec{I}$ represent the number of individuals in each state. Also assume that the epidemic can be modeled by the following generic system

$$\frac{d}{dt}\vec{S} = f(\vec{S}, \vec{I}),$$

$$\frac{d}{dt}\vec{I} = g(\vec{S}, \vec{I}),$$

Assume that $g(\vec{S}, 0) = 0$ for all $\vec{S}$; in addition, $(\vec{S}_0, \vec{0})$ is a DFE, and the initial number of infectious individuals $\vec{I}(0)$ is very small, then, initially, the dynamics of I is governed by the following linearized system

$$\frac{d}{dt}\vec{I} = \frac{\partial g}{\partial \vec{I}}(S_0, 0)\vec{I}.$$

If the DEF is unstable, then $I(t)$ grows asymptotically exponentially.

2. The exponential growth rate and the basic reproduction number

The exponential growth rate is, by itself, an important measure for the speed of spread of an infectious disease. It being zero is, like the basic reproduction number $\mathscr{R}_0 = 1$, a disease threshold. The disease can invade a population if the growth rate is positive, and cannot invade (with a few initially infectious individuals) if it is negative. In fact, it can be used to infer $\mathscr{R}_0$.There are two approaches to infer $\mathscr{R}_0$ from the exponential growth rate, a parametric one, and a non-parametric one.

2.1. The parametric approach

For the parametric approach, we need an underlying model that gives both the growth rate and $\mathscr{R}_0$.

Example 3. Consider the SIR model (1) in Example 1. Note that $(1,0,0)$ is an disease free equilibrium, representing a completely susceptible population. As we discussed above, the exponential growth rate is $\lambda = \beta - \gamma$. Note that the basic reproduction number is $\mathscr{R}_0 = \beta/\gamma$. If, for example, γ is estimated independently to λ, then,

$$\mathscr{R}_0 = \frac{\lambda}{\gamma} + 1.$$

Lets look at a more complicated example.

Example 4. Lets consider the SEIR model (3) in Example 2. The basic reproduction number is $\mathscr{R}_0 = \beta/\gamma$. To link $\mathscr{R}_0$ to the exponential growth rate

$$\lambda = \frac{-(\sigma+\gamma) + \sqrt{(\sigma-\gamma)^2 + 4\sigma\beta}}{2},$$

express β in terms of λ and substitute it into $\mathscr{R}_0$, then

$$\mathscr{R}_0 = \frac{(\lambda+\sigma)(\lambda+\gamma)}{\sigma\gamma}.$$

Thus, if the mean infectious period $1/\gamma$ and the mean latent period $1/\sigma$ can be independently estimated on λ, then $\mathscr{R}_0$ can be inferred from λ.

Typically, for an epidemic model that contains a single transmission rate β, if all other parameters can be estimated independently to the exponential growth rate λ, then λ determines β, and thus determines $\mathscr{R}_0$.

2.2. The non-parametric approach

Models can be overly simplified for mathematical tractability. For example, Both the SIR model in Example 1 and the SEIR model in Example 2 assume exponentially distributed infectious period. However, the infectious period and the latent period are mostly likely not exponential. Wallinga and Lipsitch (Wallinga & Lipsitch, 2006) developed a non-parametric method to infer the basic reproduction number from the exponential growth rate without assuming a model.

Let $\eta(a)$ be the probability that a random individual remain infectious a time units after being infected (i.e., a is the infection age); $\beta(a)$ is the rate of transmission at the infection age a. Then,

$$\tau(a) = \eta(a)\beta(a)$$

is the transmissibility of a random infectious individual at the infection age a, assuming that the whole population is susceptible. Thus,

$$\mathscr{R}_0 = \int_0^\infty \tau(a)da.$$

In addition, we assume that the population is randomly mixed, i.e., every pair of individuals have identical rate of contact. Let $c(t)dt$ be the number of new infections during the time interval $[t,t+dt]$, that is, $c(t)$ is the incidence rate, and $S(t)$ be the average susceptibility of the population, i.e., the expected susceptibility of a randomly selected individual. In addition, new infections at time t is the sum of all infections caused by infectious individuals infected a time unit ago (i.e., at time $t - a$) if they remain infectious at time t (with an infectious age a) and their contact is susceptible. That is,

$$c(t)=\int_0^\infty c(t-a)\tau(a)S(t)da,$$

and thus

$$c(t)=S(t)\int_0^\infty c(t-a)\eta(a)da.$$

To compute $\mathscr{R}_0$, we need to normalize $\tau(a)$ as a probability density function,

$$w(a)=\frac{\tau(a)}{\int_0^\infty \tau(s)ds}=\frac{\tau(a)}{\mathscr{R}_0}.$$

Note that $w(a)da$ is the probability that a secondary infection occurs during the infection age interval $[a,a+da]$. That is, $w(a)$ is the probability density function of the generation time, i.e., the time from being infected to generate a secondary infection. This generation time is also called the serial interval. With the serial interval distribution $w(t)$,

$$c(t)=\mathscr{R}_0S(t)\int_0^\infty c(t-a)w(a)da. \tag{4}$$

This means that the $c(t)$ is only determined by $\mathscr{R}_0$, $w(t)$ and $S(t)$. At the beginning of an epidemic, where the epidemic grows exponentially (with an exponential growth rate λ), $S(t)\approx 1$ and $c(t)=c_0e^{\lambda t}$ where c_0 is the initial number of cases at time $t=0$. Thus,

$$e^{\lambda t}=\mathscr{R}_0\int_0^\infty e^{\lambda(t-a)}w(a)da,$$

that is,

$$\mathscr{R}_0=\frac{1}{\int_0^\infty e^{-\lambda a}w(a)da}=\frac{1}{M(-\lambda)}, \tag{5}$$

where $M(x)=\int_0^\infty e^{xa}w(a)da$ is the moment generating function of the serial time distribution $w(a)$.

Equation (5) links the exponential growth rate to the basic reproduction number though the serial interval distribution only. That is, if we can estimate the serial interval distribution and the exponential growth rate independently, that we can infer the basic reproduction number.

Note that the serial interval distribution $w(t)$ can be estimated independently to the exponential growth rate. For example, it can be estimated empirically using contact tracing. Alternatively, one can also assume an epidemic model. Here we discuss a few simple examples.

Example 5. Consider an SIR model. Let $F(a)$ be the cumulative distribution function of the infectious period, and a constant transmission rate β. The probability that an infected individual remains infectious a time units after being infected is

$$\eta(a)=1-F(a),$$

and thus the transmissibility is

$$\tau(a)=\beta[1-F(a)],$$

and the serial interval distribution is

$$w(a)=\frac{\tau(a)}{\int_0^\infty \tau(t)dt}=\frac{1-F(a)}{\int_0^\infty 1-F(a)dt}=\frac{1-F(a)}{\mu},$$

where μ is the mean infectious period. For the special case that the infectious period is exponentially distributed with a rate γ, i.e., $F(a)=1-e^{-\gamma a}$, this model becomes Model (1). Then the density function of serial interval distribution is

$$w(a) = \frac{1 - F(a)}{1/\gamma} = \gamma e^{-\gamma a}$$

which is identical to the density function of infectious period distribution. The moment generating function is

$$M(x) = \frac{\gamma}{\gamma - x},$$

Note that the exponential growth rate is $\lambda = \beta - \gamma$, then

$$\mathscr{R}_0 = \frac{1}{M(-\lambda)} = \frac{\gamma + \lambda}{\gamma} = \frac{\beta}{\gamma}.$$

Lets consider a more complex example with multiple infected states.

Example 6. Consider an SEIR model with a constant transmission rate β. Let $F(a)$ and $G(a)$ be the cumulative distribution functions of the infectious period and the latent period, respectively. Given the latent period $T_L = \ell \le a$, the probability that an infectious individual is infectious a time units after being infected is $1 - F(a - \ell)$.Thus,

$$\eta(a) = \int_0^a 1 - F(a - \ell) dG(\ell).$$

Hence, the serial interval distribution is

$$w(a) = \frac{\int_0^a [1 - F(a - \ell)]G'(\ell)d\ell}{\int_0^\infty \int_0^a [1 - F(a - \ell)]G'(\ell)d\ell da}.$$

For the special case that the latent period is exponentially distributed with a rate σ (i.e., $F(a) = 1 - e^{-\gamma a}$) and the latent period is exponentially distributed with a rate σ (i.e., $G(a) = 1 - e^{-\sigma a}$), this model becomes Model (3), and

$$w(a) = \gamma\sigma e^{-\gamma a} \int_0^a e^{(\gamma - \sigma)s} ds = (\gamma e^{-\gamma a}) * (\sigma e^{-\sigma a}).$$

That is, if both distributions are exponential, the serial interval distribution is the convolution of the latent period distribution and the infectious period distribution. In this case, the basic reproduction number is

$$\mathscr{R}_0 = \frac{1}{M(-\lambda)} = \frac{1}{M_I(-\lambda)M_L(-\lambda)} = \frac{(\lambda + \gamma)(\lambda + \sigma)}{\gamma\sigma},$$

where $M_I(x)$ and $M_L(x)$ are the moment generating functions of the infectious period and latent period, respectively.

Remark

In Equation (4), $\mathscr{R}(t) = \mathscr{R}_0 S(t)$ is the reproduction number, and thus this equation can be used to estimate the production number at any time t during the epidemic given the incidence curve $c(t)$, namely,

$$\mathscr{R}(t) = \frac{c(t)}{\int_0^\infty c(t - a)w(a)da}.$$

This is similar to, but different from, the nonparametric method developed by Wallingua and Teunis (Wallinga & Teunis, 2004).

3. Least squares estimation

The least squares method is one of the most commonly used methods for parameter estimation in mathematical biology. This method is in fact a mathematical method. For a family of curves $f(t;\ \overrightarrow{\theta})$, where $\overrightarrow{\theta} \in \mathbb{R}^m$ is a vector of parameters of the family, this method finds the curve $f(t;\widehat{\theta})$ in the family that minimizes the distance between the curve and a set of points $\{(t_i, x_i)\}_{i=0}^{n-1}$. Let $\overrightarrow{x} = (x_0, \ldots, x_{n-1})$, and $\overrightarrow{f}(\overrightarrow{\theta}) = (f(t_0;\ \overrightarrow{\theta}), \ldots, f(t_{n-1};\ \overrightarrow{\theta}))$, and $\overrightarrow{x}$ be the Euclidean norm in $\mathbb{R}^n$, then the mathematical formulation of the least squares method is

$$\widehat{\theta} = \underset{\overrightarrow{\theta}}{argmin} \| \overrightarrow{f}(\overrightarrow{\theta}) - \overrightarrow{x} \|^2 = \underset{\overrightarrow{\theta}}{argmin} \sum_{i=0}^{n-1} [f(t_i;\ \overrightarrow{\theta}) - x_i]^2, \tag{6}$$

where *argmin* gives the parameter $\overrightarrow{\theta}$ that minimizes the objective function. For our purpose, the observations $\{(t_i, x_i)\}_{i=0}^{n-1}$ is the epidemic curve, i.e., x_0 is the number of initially observed cases, and x_i is the number of new cases during the time interval $(t_{i-1}, t_1]$. We aim to find an exponential function $f(t; c_0, \lambda) = c_0 e^{\lambda t}$ that minimizes its distance to the epidemic curve, i.e., the parameters $\theta = (c_0, \lambda)$. There are two commonly use methods to estimate the exponential growth rate λ:

1. Nonlinear least square to fit to $f(t; c_0, \lambda) = c_0 e^{\lambda t}$ directly;
2. Linear least square to fit $\{(t_i, \ln x_i)\}$ to $\ln f(t; c_0, \lambda) = \ln c_0 + \lambda t$.

The nonlinear least squares method does not have an analytic solution. Numerical optimization is needed to solve the minimization problem (6). The linear least square method has an analytic solution: Let $\ell_0 = \ln c_0$, then the least squares problem becomes

$$(\ell_0, \lambda) = \underset{(\ell_0, \lambda)}{argmin} \sum_{i=0}^{n-1} (\ell_0 + \lambda t_i - \ln x_i)^2.$$

The objective function is a quadratic function of ℓ_0 and λ, thus, the minimum is achieved at $(\widehat{\ell}_0, \widehat{\lambda})$ that satisfies

$$\frac{\partial s}{\partial \ell_0}\Big|_{\substack{\ell_0=\hat{l}_0 \\ \lambda\ =\widehat{\lambda}}\ \widehat{\ }\ell_0=l_0} = \sum_{i=0}^{n-1} 2\left(\widehat{\ell}_0 + \widehat{\lambda} t_i - \ln x_i\right) = 0,$$

$$\frac{\partial s}{\partial \lambda}\Big|_{\substack{\ell_0=\hat{l}_0 \\ \lambda\ =\widehat{\lambda}}\ \widehat{\ }\ell_0=l_0} = \sum_{i=0}^{n-1} 2t_i\left(\widehat{\ell}_0 + \widehat{\lambda} t_i - \ln x_i\right) = 0.$$

Let $\langle y_i \rangle = \frac{1}{n}\sum_{i=0}^{n-1} y_i$, which represents the average of any sequence $\{y_i\}_{i=0}^{n}$, then,

$$\begin{bmatrix} 1 & \langle t_i \rangle \\ \langle t_i \rangle & \langle t_i^2 \rangle \end{bmatrix} \begin{bmatrix} \widehat{\ell}_0 \\ \widehat{\lambda} \end{bmatrix} = \begin{bmatrix} \langle \ln x_i \rangle \\ \langle t_i \ln x_i \rangle \end{bmatrix},$$

and thus the best fit exponential growth rate ls

$$\widehat{\lambda} = \frac{\langle t_i \ln x_0 \rangle - \langle t_i \rangle \langle \ln x_i \rangle}{\langle t_i^2 \rangle - \langle t_i \rangle^2}.$$

Do these two methods yield the same answer? To compare, we simulate an epidemic curve of the stochastic SEIR model in Example 2, using the Gillespie method (Gillespie, 1976). The simulated daily cases (number of individuals showing symptom on a day) are then aggregated into weekly cases. Then, we use both methods to fit an exponential curve to the simulated epidemic curve. The simulated epidemic curve and the fitting results are shown in Fig. 3. This exercise illustrates a challenge of fitting an exponential model to an epidemic curve: how to determine the time period to fit the exponential model. The exponential growth rate of an SEIR model decreases with time as the susceptible population decreases. In Fig. 3, The epidemic curve peaks in week 13. We choose a sequence of nested fitting windows starting in the first week and ending in a week w for $w = 3, 4, \ldots, 13$. The SEIR model has an asymptotic exponential growth, so the fitted exponential growth rate is not monotonic near the beginning of the epidemic. For larger fitting windows, both methods give an exponential growth rate that decreases with the length of the fitting window. We need more data points to reduce the influence of the stochasticity. However, using more data points also risks of obtaining an estimate that deviates too much from the true exponential growth rate. There is no reliable method to choose a proper fitting window.

Fig. 3 also shows that the linear and nonlinear least squares methods may not yield the same estimate. This is because of a major limitation of both least squares methods: they implicitly assume that the deviations $|x_i - f(t_i;\ \overrightarrow{\theta})|$ carry identical

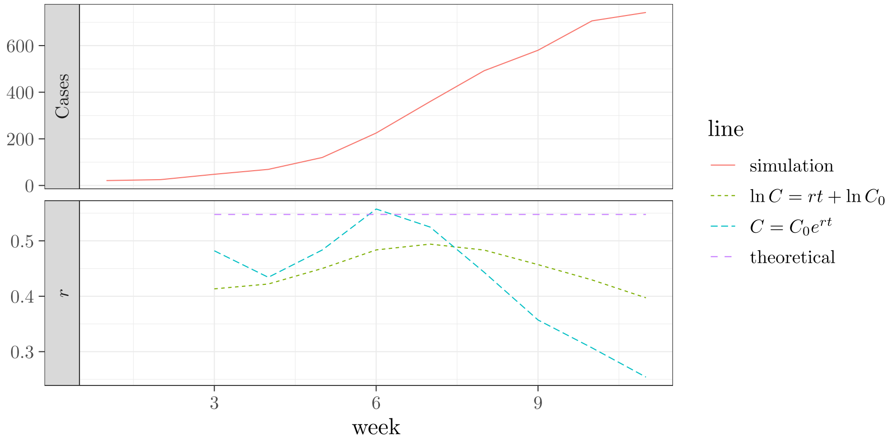


Fig. 3. The simulated SEIR epidemic curve (upper) and the fitted exponential growth rate as a function of the end of the fitting window (lower). The epidemic curve is simulated stochastically from the SEIR model in Example 2 using the Gillespie method (Gillespie, 1976) with the parameters $\beta = 0.3$, $\sigma = 1$, $\gamma = 0.2$, $E_0 = 10$, $S_0 = 9,990$. $I_0 = R_0 = 0$. The rates have a time unit of a day. The daily cases are then aggregated by week. The data points are taken at times $t_i = i$, $i = 0, 1, 2, \ldots 13$ weeks. The theoretical exponential growth rate is $\lambda = 0.547$ per week.

weights. With the nonlinear method, later data points (at larger times) deviate more from the exponential curve than the earlier data points, because the exponential growth slows down with time. Thus, the method is more biased to the later data points. With the linear method, the deviations in $\ln x_i$ are more even than in x_i, and thus the linear method is less biased to the later data points than the nonlinear method does.

The least squares method, as mentioned above, is a mathematical problem. It does not explicitly assume any error distributions, and thus cannot give us statistical information about the inference. For example, if we use two slightly different fitting windows and get two slightly different estimates, is the difference of the two estimates statistically significant? Such a question cannot easily be answered by the least squares method. Interestingly, the least squares methods make many implicit assumptions to the deviations. We have mentioned the implicit equal-weight assumption above. It also implicitly assumes that the order of the observations does not matter, and that positive and negative deviations are equivalent. Thus, they implicitly assume that the deviations are independently identically and symmetrically distributed. In statistics, the least squares method is commonly used in linear and nonlinear regression with an addition assumption that the errors are independently and identically normally distributed. However, these assumption on the errors may not be appropriate. For example, the new cases at time $t+1$ may be infected by those who are infected at time t. Thus, the number of new cases at different times may not be independent. Also, the number of cases is a counting variable, and thus its mean and variance may be closely related, meaning that the error may not be identically normally distributed. In the next section, we address some of these problems using the maximum likelihood method.

4. Maximum likelihood estimation

The maximum likelihood method is a commonly used statistical method for parameter inference; see, e.g., [(Bolker, 2008), p.170]. It relies on a “likelihood function” $L(\overrightarrow{\theta})$ where $\overrightarrow{\theta}$ is the vector of parameters. The likelihood function is a function proportional to the conditional probability of observing the data points $\{(t_i, x_i)\}_{i=0}^{n-1}$ given the parameters $\overrightarrow{\theta}$, i.e.,

$$L(\overrightarrow{\theta}) \propto P\left(\{(t_i, d_i)\}_{i=0}^{n-1} \middle| \overrightarrow{\theta}\right).$$

We choose the parameter values that maximize the likelihood, i.e.,

$$\widehat{\theta} = \underset{\overrightarrow{\theta}}{\mathrm{argmin}} L(\overrightarrow{\theta}).$$

To construct the likelihood function we need to make assumptions on the error distribution. There are two types of error: the process error and the observation error. The observation error is the error in the observation process. For example, most people with influenza do not go to see a doctor, and thus there is no record of these cases, resulting in an under-reporting of the number influenza cases. Also, many influenza related deaths are caused by complications such as pneumonia, and influenza may not be recorded as the cause. Typos, miscommunication, etc, can all result in observation errors. The process error originates from the stochasticity of the system that is independent to observation. For example, the disease dynamics is intrinsically stochastic. The time that an infectious individual recovers, and the time that a susceptible individual is infected, are all random variables that affects the number of new infections at any time, even if we eliminate all observation errors. These two types of errors have very different nature, and thus need very different assumptions. For example, it is reasonable to assume that observation errors are independent to each other, but process errors at a later time are commonly dependent on the process errors at earlier times.

4.1. Case 1: process errors are negligible

If observation errors are large and process errors are negligible, then we assume that the random variable X_i corresponding to the observation x_i is independently distributed with a probability mass function $p_i(k; \overrightarrow{\theta})$ where k is the values that X_i can take. Then, the likelihood function is

$$L(\overrightarrow{\theta}) = \prod\nolimits_{i=0}^{n-1} p_i(d_i; \overrightarrow{\theta}).$$

The maximization of this likelihood function rarely has an analytic solution, and commonly needs to be solved numerically. Note that each factor (probability) can be very small, and thus the product may be very difficult to minimize numerically because of rounding errors (from the binary representation of real numbers in computers). It is a common practice to maximize the log-likelihood function

$$\ell(\overrightarrow{\theta}) = \ln L(\overrightarrow{\theta}) = \sum_{i=0}^{n} \ln p_i(d_i; \overrightarrow{\theta}).$$

For example, we assume that the number of cases $x(t_i)$ at time t_i is independently Poisson distributed with mean $\mu_i = c_0 e^{\lambda t_i}$. Then, the log-likelihood function

$$\ell(c_0, \lambda) = \sum_{i=0}^{n} \ln \frac{e^{-\mu_i} \mu_i^{x_i}}{x_i!} = \sum_{i=0}^{n-1} - \mu_i + x_i \ln\mu_i - \ln x_i!.$$

Note that the observed cases x_i are constants, and thus the last term can be ignored for maximization. Thus,

$$(\widehat{c}_0, \widehat{\lambda}) = \underset{(c_0,\lambda)}{\operatorname{argmax}} \sum_{i=0}^{n-1} - \mu_i + x_i \ln\mu_i = \underset{(c_0,\lambda)}{\operatorname{argmax}} \sum_{i=0}^{n-1} - c_0 e^{\lambda t_i} + x_i \ln c_0 + \lambda x_i t_i.$$

This maximization problem can only be solved numerically.

We choose Poisson distribution because its simple form greatly simplifies the log-likelihood function. In addition, it does not introduce more parameters, which is valuable to avoid over-fitting when the number of data points available is small. If the process error is not completely negligible, then choosing an overly dispersed distribution, such as the negative binomial distribution may be desirable. A negative binomial distribution has two parameters, the success probability $q \geq 0$ and the shape parameter $r > 0$. For simplicity, we assume that the shape parameter r is the same at each time t_i, and will; be estimated together with the model parameters $\overrightarrow{\theta}$; but q depend on t_i. The probability mass function is

$$p_i(k; q_i, r) = \frac{\Gamma(k+r)}{k!\Gamma(r)} q_i^k (1 - q_i)^r,$$

with the mean

$$\mu_i = \frac{q_i r}{1 - q_i} = c_0 e^{\lambda t_i}.$$

Thus,

$$q_i = \frac{c_0 e^{\lambda t_i}}{r + c_0 e^{\lambda t_i}},$$

and the log-likelihood function is

$$\ell(c_0, \lambda, r) = \sum_{i=0}^{n-1} \ln \frac{\Gamma(x_i + r)}{x_i! \Gamma(r)} \frac{c_0^{x_i} e^{\lambda x_i t_i} r^r}{(r + c_0 e^{\lambda t_i})^{x_i + r}}$$

$$= \sum_{i=0}^{n-1} \ln\Gamma(x_i + r) - \ln\Gamma(r) + x_i c_0 + \lambda x_i t_i + r \ln r - (x_i + r)\ln\left(r + c_0 e^{\lambda t_i}\right) - \ln x_i!.$$

Again, the last term can be ignored for the optimization problem. In addition, there is a constraint $r > 0$.

4.2. Case 2: observation errors are negligible

If process errors are large and observation errors are negligible, then we cannot assume that the observed values X_{i+1} and X_i are independent to each other. Instead, for all $i = 0, 1, \ldots, n-2$, we compute the probability mass function of X_{i+1} given $\{X_j = x_j\}_{j=0}^{i}$, namely, $q_{i+1}(k; \overrightarrow{\theta} \,|\, \{x_j\}_{j=0}^{i})$. Then, the likelihood function is

$$L(\overrightarrow{\theta}) = P\left(\{x_i\}_{i=0}^{n-1} \middle| \overrightarrow{\theta}\right)$$

$$= P\left(x_{n-1} \middle| \{x_i\}_{i=0}^{n-2}, \overrightarrow{\theta}\right) P\left(\{x_i\}_{i=0}^{n-2}, \overrightarrow{\theta}\right)$$

$$= \prod_{i=1}^{n-1} q_i\left(x_i; \overrightarrow{\theta} \middle| \{x_j\}_{j=0}^{i-1}\right).$$

For simplicity, assume that X_{i+1} is Poisson distribution with mean $\mu_{i+1} = X_i e^{\lambda(t_{i+1} - t_i)}$. Note that, since we assumed no observation error, the initial condition $c_0 = x_0$ is exact, and thus there is a single parameter λ for the model. Thus,

$$q_{i+1}\left(k; \overrightarrow{\theta} \middle| \{x_j\}_{j=0}^{i}\right) = \frac{e^{\mu_{i+1}} \mu_{i+1}^{k}}{k!},$$

and thus the log-likelihood function is

$$l(\overrightarrow{\theta}) = \sum_{i=0}^{n-1} \ln \frac{e^{\mu_i} \mu_i^{x_i}}{x_i!}$$

$$= \sum_{i=0}^{n-1} \mu_i + x_i \ln\mu_i - \ln x_i!$$

$$= \sum_{i=0}^{n-1} x_{i-1} e^{\lambda(t_i - t_{i-1})} + x_i \lambda(t_i - t_{i-1}) + x_i \ln x_i - \ln x_i!.$$

Again, the last two terms can be ignored in maximization because they are constants. Thus,

$$\lambda = \operatorname*{argmax}_{\lambda} x_{i-1} e^{\lambda(t_i - t_{i-1})} + (t_i - t_{i-1}) x_i \lambda.$$

4.3. Case 3: consider both type of errors together

It is much harder to formulate the likelihood function if process errors and observation errors must both be considered. We can simplify the problem by ignoring the process error and use an overly dispersed observation error distribution as a compensation. Note that this simplification mainly affects the confidence intervals.

4.4. Confidence intervals

The maximum likelihood method gives a point estimate, i.e., one set of parameter values that makes it mostly likely to observe the data. However, it is not clear how close the point estimates are to the real values. To answer this question we use an interval estimate, commonly known as a confidence interval. A confidence interval with a confidence level α is an interval that has a probability α that contains the true parameter value. A commonly used confidence level is 95%, which originates from a normal distribution. If a random variable X is normally distributed with a mean μ and a standard deviation σ, then the probability that $X \in [\mu - 2\sigma, \mu + 2\sigma]$ is 95%.

The confidence interval can be estimated using the likelihood ratio test [(Bolker, 2008), p.192]. Let $\widehat{\overrightarrow{\theta}}$ be the point estimate of the parameters. A value λ_0 is in the 95% confidence interval is equivalent to accepting with 95% probability that λ_0 is a possible growth rate. To determine this we fit a nested model by fixing the growth rate $\lambda = \lambda_0$, suppose its point estimate is $\widehat{\theta}_0$. We then compute the likelihood ratio

$$\Lambda = \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})} \; .$$

The Wilks' theorem (Wilks, 1938) guarantees that, as the sample size becomes large, the statistics $-2\ln\Lambda = 2[\ell(\widehat{\theta}) - \ell(\widehat{\theta}_0)]$ is χ^2 distributed with a degree of freedom 1. We thus can compare $-2\ln\Lambda$ with the 95% quantile of the χ^2 distribution and determine if λ_0 should be in the confidence interval or not. We can thus perform a linear search on both sides of the point estimate to determine the boundary of the confidence interval.

5. Mechanistic and phenomenological models

We still have not addressed the problem of choosing a fitting window for an exponential model. Recall that the challenge arises because the exponential growth rate of an epidemic decreases with time. Instead of finding heuristic conditions for choosing the fitting window, we circumvent this problem by incorporating the decrease of the exponential growth rate into our model. We have two choices, using either a mechanistic model such as an SIR or SEIR model, or a phenomenological model.

5.1. Mechanistic models

Naturally, if we know that a mechanistic model is a good description of the disease dynamics, fitting such a model to the epidemic curve is a good option (see, e.g., (Chowell, Ammon, Hengartner, & Hyman, 2006; Pourabbas, d'Onofrio, & Rafanelli, 2001),). We use an SIR model as an example. For simplicity, we assume that the process error is negligible, and the incidence rate is Poisson distributed with a mean $C(t)$ given by an SIR model ($C(t) = \beta SIN$ where N is the population size). To construct the log-likelihood function, we need to calculate $C(t)$, i.e., numerically solve the SIR model. To do so, we need the transmission rate β. the recovery rate γ, the initial fraction of infectious individuals $I(0) = I_0$ (with the assumption that $R(0) = 0$, $S(0) = 1 - I_0$, and thus I_0 determines the initial conditions), in addition to the population size N. Thus, the parameters of the model is $\overrightarrow{\theta} = (\beta, \gamma, I_0, N)$. Thus the log-likelihood function is (ignoring the constant terms)

$$\ell(\beta, \gamma, I_0, N) = \sum_{i=0}^{n-1} -c(t_i) + x_i \ln c(t_i) \; ,$$

where the number of new cases $c(t_i)$ in the time interval $[t_i, t_{i+1}]$ is

$$c(t_i) = S(t_{i+1}) - S(t_i) \; ,$$

and $S(t_i)$ is solved numerically from the SIR model. Thus, ℓ implicitly depend on β, γ and I_0 through $S(t)$.

One draw back using such a mechanistic model is its high computational cost, since each evaluation of the log-likelihood function requires solving the model numerically, and numerical optimization algorithms can be very hungry on function evaluations, especially if the algorithm depends on numerical differentiation.

Another draw back is that these mechanistic models can be overly simplified, and may not be a good approximation to the real disease dynamics. For example, for seasonal influenza, due to the fast evolution of the influenza virus, individuals have different history of infection, and thus have different susceptibility to a new strain. Yet simple SIR and SEIR models assume a population with a homogeneous susceptibility. Thus using a simple SIR to fit to an influenza epidemic may be an over simplification. However, realistic mechanistic models can be overly complicated, and involve too many parameters that are at best difficult to estimate. For example, a multi-group SIR model depends on a contact matrix consisting of transmission rates between groups, which contains a large number of parameters if the model uses many groups.

5.2. Phenomenological models

If all we need to estimate is the exponential growth rate, we only need a model that describes the exponential growth that gradually slows down. Most cumulative epidemic curves grow exponentially initially, and then saturates at the final epidemic size. A simple phenomenological model can be used to describe the shape of the cumulative epidemic curve, but the model itself may not have realistic biological meaning. However, if simple mechanistic models cannot faithfully describe the epidemic process, using a simple phenomenological model with an analytical formula may be a better choice, at least numerically, because repetitively solving a system differential equations numerically, and differentiating the log-likelihood function numerically, can both be avoided with the analytical formula. Here we discuss some examples for such models.

Logistic model

The logistic model is the simplest model that shows an initial exponential growth followed a gradual slowing down and a saturation. The cumulative incidences $C(t)$ (the total number of cases by time t) can be approximated by

$$\frac{d}{dt}C(t) = rC(t)\left(1 - \frac{C(t)}{K}\right).$$

where r is the exponential growth rate, and $K = \lim_{t\to\infty}C(t)$. Let $C_0 = C(0)$, its solution is

$$C(t) = \frac{KC_0}{C_0 + (K - C_0)e^{-rt}}, \tag{7}$$

The new cases $c(t_i)$ in a time period $[t_i, t_{i+1}]$ is thus

$$c(t_i) = C(t_{i+1}) - C(t_i)\ . \tag{8}$$

The model parameters are $\overrightarrow{\theta} = (r, K, C_0)$. Note that it is less than the number of parameters of the simplest mechanistic model (i.e., the SIR model).

Richards model

The logistic model has a fixed rate of slowing down of the exponential growth rate. To be more flexible, we can use the Richards model (Richards, 1959) for the cumulative incidence curve. The Richards model, also called the power law logistic model, can be written as

$$\frac{d}{dt}C(t) = rC(t)\left[1 - \left(\frac{C(t)}{K}\right)^{\alpha}\right],$$

where $\alpha$is the parameter that controls the steepness of the curve. Note that the logistic model is a special case with $\alpha = 1$. Its solution is

$$C(t) = \frac{KC_0}{\left[C_0^{\alpha} + \left(K^{\alpha} - C_0^{\alpha}\right)e^{-\frac{rtK^{\alpha}}{K^{\alpha} - C_0^{\alpha}}}\right]^{1/\alpha}}.$$

The new cases $c(t_i)$ in a time period $[t_i, t_{i+1}]$ is also given by (8). The parameters are $\overrightarrow{\theta} = (r, K, C_0, \alpha)$.

5.3. Comparison of the models

To compare the performance of both the SIR model and the phenomenological models, we fit these models to the stochastically simulated SEIR epidemic curve of weekly cases that we introduced in Section 3 (Fig. 3).

We assume that the process error is negligible, and the observations are Poisson distributed about the mean that is given by the corresponding models. We use the maximum likelihood method. The results are shown in Fig. 4. The predictions of the exponential model, as discussed before, quickly decreases as more data points are used. Both the logistic model and the Richards model give robust estimates with fitting windows ending up to the peak of the epidemic. The SIR model gives a robust estimate for all fitting windows up to the whole epidemic curve.

Thus, the SIR model is a good model to use to fit the exponential growth rate, even if it may not be the correct mechanistic model. (e.g., it ignores the latent period in this example). It requires more computational power, because the epidemic curve lacks an analytic formula, and needs to be numerically solved from a system of ordinary differential equations. The logistic model and the Richards model can be used for all data points up to the peak of the epidemic.

5.4. Coverage probabilities

Fig. 4 also show that the SIR model and the logistic model give the narrowest confidence intervals. However, narrower confidence intervals may not be desirable if it has a large chance that it does not contain the true value. Due to errors, especially process errors, each realization of the underlying stochastic epidemic process yields a different epidemic curve. These epidemic curves may exhibit different exponential growth rates even if the underlying parameter values are the same. An observed epidemic curve is just a single realization of the epidemic process. Does the estimated confidence intervals contain the theoretical exponential growth rate of the epidemic process? This question is answered by the "coverage probability", which is the probability that the confidence interval contains the true value. If the confidence interval properly considers all sources of stochasticity, then the coverage probability should be equal to its confidence level.

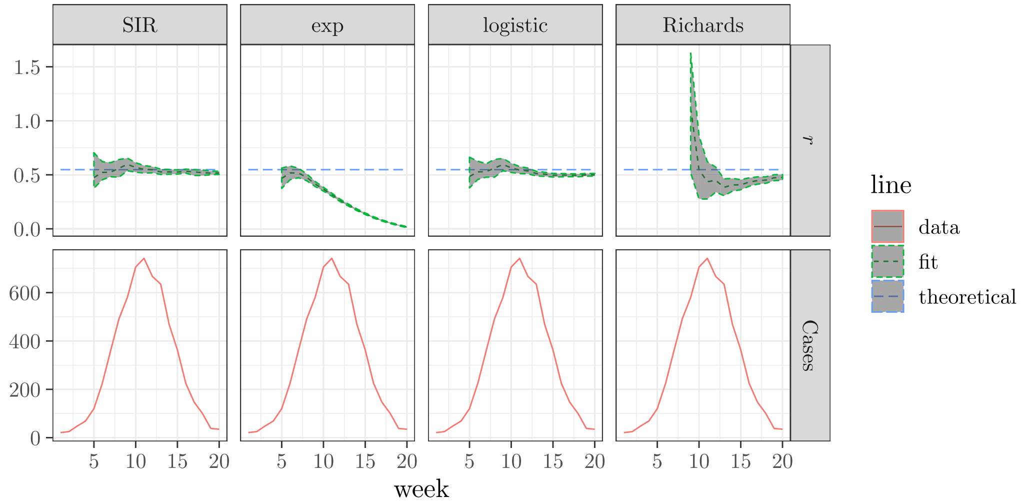


Fig. 4. The comparison of the results of fitting the SIR, exponential, logistic, and Richards models to a simulated weekly incidence curve, as a function of the end point of the fitting window (upper). The epidemic curve (lower) is shown as a reference. The epidemic curve and the theoretical exponential growth rate are the same as Fig. 3 s.

To illustrate this, we numerically compute the coverage of the confidence intervals by simulating the SEIR model 400 times and compute confident interval of the exponential growth rate for each realization, and compute the fraction of the confident intervals containing the theoretical value $\lambda = 0.537$. The results is summarized in below:

	logistic model	Richards model
coverage probability	43%	65%

That is, even though the logistic model gives a narrow confidence interval, its coverage probability is low. The coverage probability of the confidence interval given by the Richards model is also significantly lower than the confidence level. This is indeed caused by treating process errors as observation errors. If there is under reporting, that is, only a fraction p of the cases can be observed, then the observation error becomes larger as p decreases (i.e., more under reporting). The coverage will become larger as a result. For example, the case fatality ratio of the 1918 pandemic influenza is about 2% (Frost, 1920). Thus, the mortality curve can be treated as the epidemic curve with a large under reporting ratio, and thus the observation error dominates. In this case ignoring the process error is appropriate.

Acknowledgements

This research is partially supported by a Natural Sciences and Engineering Research Council Canada discovery grant, and National Natural Science Foundation of China (No.11771075).

References

Bolker, B. M. (2008). *Ecological models and data in R*. Princeton: Princeton University Press.

Center for Disease Control. (2016). Ebola: Case counts. *Tech. rep.* https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html.

Chowell, G., Ammon, C. E., Hengartner, N. W., & Hyman, J. M. (2006). Transmission dynamics of the great influenza pandemic of 1918 in geneva, Switzerland: Assessing the effects of hypothetical interventions. *Journal of Theoretical Biology, 241*, 193–204.

Frost, W. H. (1920). Statistics of influenza morbidity. with special reference to certain factors in case incidence and case-fatality. *Public Health Reports, 35*, 584–597.

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics, 22*, 403–434.

Ma, J., Dushoff, J., Bolker, B. M., & Earn, D. J. D. (2013). Estimating initial epidemic growth rates. *Bulletin of Mathematical Biology, 76*, 245–260.

Pourabbas, E., d'Onofrio, A., & Rafanelli, M. (2001). A method to estimate the incidence of communicable diseases under seasonal fluctuations with application to cholera. *Applied Mathematics and Computation, 118*, 161–174.

Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany, 10*, 290–300.

Wallinga, J., & Lipsitch, M. (2006). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences, 274*, 599–604.

Wallinga, J., & Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology, 160*, 509–516.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics, 9*, 60–62.