Estimating the Exponential Growth Rate and $R_0$

Junling Ma
Department of Mathematics and Statistics,
University of Victoria

May 23, 2012
Introduction

Daily pneumonia and influenza (P&I) deaths of 1918 pandemic influenza in Philadelphia.
Re-examine 1918 Daily Philadelphia P&I Deaths

▶ An exponential growth phase
▶ Given infectious period and latent period, this rate implies $R_0$

Junling Ma
Department of Mathematics and Statistics, University of Victoria

Estimating the Exponential Growth Rate and $R_0$
What We Will Learn

- Estimate the exponential growth rate
- Fit (phenomenological or mechanistic) models to data
- Estimate $R_0$ from the exponential growth rate
The Exponential Growth Phase

- The 1918 pandemic epidemic curve, and most others, show an initial exponential growth phase,
- That is, during the initial growth phase, the epidemic curve can be modeled as
  \[ X(t) = X(0)e^{\lambda t}, \]
  where \( \lambda \) is the exponential growth rate, \( X(0) \) is the initial condition.
- So, \( \ln X(t) \) and the time \( t \) have a linear relationship during the initial growth phase
  \[ \ln X(t) = \ln X(0) + \lambda t. \]
- The exp growth rate measures how fast the disease spreads...
Example: 1665 Great Plague Deaths in London

- Exponential growth rate decreases around week 25?
Example: 1665 Great Plague All-cause Deaths in London

- The decrease of the exponential growth rate in plague deaths may be caused by under reporting.
Theoretical Exponential Growth Rate: SIR Model

\[ S' = -\frac{\beta}{N} SI, \quad I' = \frac{\beta}{N} SI - \gamma I. \]

- Near the disease free equilibrium (DFE) \((N, 0)\)
  \[ I' = (\beta - \gamma)I \]

- This is a linear ODE, with an exponential solutions
  \[ I(t) = I(0)e^{(\beta - \gamma)t} \]

- So, the exponential growth rate is \(\lambda = \beta - \gamma\).

- What is the growth rate of the incidence curve \(X(t) = \beta SI?\)
Theoretical Growth Rate: SEIR Model

\[ S' = -\frac{\beta}{N}SI , \quad E' = \frac{\beta}{N}SI - \sigma E . \quad I' = \sigma E - \gamma I . \]

- Near the disease free equilibrium (DFE) \( (N, 0, 0) \)

\[ \frac{d}{dt} \begin{bmatrix} E \\ I \end{bmatrix} = \begin{bmatrix} -\sigma & \beta \\ \sigma & -\gamma \end{bmatrix} \begin{bmatrix} E \\ I \end{bmatrix} = J \begin{bmatrix} E \\ I \end{bmatrix} \]

- The exponential growth rate is

\[ \lambda = \rho(J) = \frac{1}{2} \left( \lambda + \gamma + \sqrt{(\sigma - \gamma)^2 + 4\beta\gamma} \right) \]
Theoretical Growth Rate: General Case

Assume that a disease can be modeled with

- Susceptible classes $S \in \mathbb{R}^m$ and infected classes $I \in \mathbb{R}^n$
- parameters $\theta \in \mathbb{R}^p$.
- Assume a disease free equilibrium (DFE) ($S = S^*, I = 0$).

$$S' = f(S, I; \theta), \quad I' = g(S, I; \theta), \quad \text{where} \quad \frac{\partial}{\partial S} g(S^*, 0) = 0$$

- Linearize about the DFE ($S^*, 0$):

$$I' = \frac{\partial g}{\partial I}(S^*, 0; \theta) I.$$

- The exponential growth rate

$$\lambda = \rho \left( \frac{\partial g}{\partial I}(S^*, 0; \theta) \right)$$
Fitting an Exponential Curve

Model

\[ x(t) - x(0)e^{\lambda t} \]

Naive methods that have been widely used:

- Least square and linear regression
- Poisson regression
- Negative binomial regression

These methods

- assume a mean that can be described by a deterministic model
- only consider observation errors around the deterministic model
- ignore the process errors are completely ignored
Point Estimates and Confidence Intervals

- The best estimate for \((\lambda, x_0)\) is called a *point* estimate.
- A 95% confidence interval (CI) \((a, b)\) for \(\lambda\) is an *interval* estimate that satisfies
  \[
  \text{Prob}\{\lambda \in (a, b)\} = 95\%
  \]
- 95% is called the confidence level. Other examples, 99% CI
- Infinitely many CI with the same confidence level (95%)
- Wider CIs means the true parameter value may differ more widely from the point estimate
- E.g.: The 1918 pandemic influenza (fall wave) has \(R_0 = 1.86\) with 95% CI \((1.82, 1.90)\) (Chowell et al Proc B 2008).
Fitting an Exponential Curve

**Linear Regression**

\[ x(t) = x(t)e^{\lambda t} \Rightarrow \ln x(t) = \ln x(0) + \lambda t \]

Commonly use the least square method:

- For a data set \((t_i, x_i)\), minimize
  \[
  F(\lambda, x_0) = \sum_{t=1}^{n} (\ln x(t_i) - \ln x_i)^2
  \]

- Confidence intervals:
  - Assume that \(\ln x_i\) are normally distributed,
    - i.e., \(x_i\) are log-normally distributed
    - Then \((\lambda, x_0)\) are joint normal
    - The covariance matrix is \((D^2F)^{-1}\).
  - If \(x_i\) is not log-normal, not an easy problem.
Poisson Regression

- For an epidemic curve \((t_i, x_i), x_i\) usually not \(\sim\) log-normal.
- If infection events have exponentially distributed waiting time, \(x_i\) are *Poisson* distributed.
- Poisson regression for these type of data, which is a maximum likelihood method.
  - A likelihood function is the probability of observing the data with a given set of parameters

\[
L(\{x_i\}_{i=1}^n | \lambda, x_0) = \prod_{i=1}^n \text{Prob}(x_i | \lambda, x_0) = \prod_{i=1}^n \frac{E[x_i]^x_i e^{-E[x_i]}}{x_i!}
\]

\[
= \prod_{i=1}^n \frac{x(t_i)^{x_i} e^{-x(t_i)}}{x_i!} = \prod_{i=1}^n \frac{x_0^{x_i} e^{\lambda t_i x_i - x_0 \exp(\lambda t_i)}}{x_i!}.
\]

- Find the parameters \(\lambda, x_0\) that maximize \(L\)
Poisson Regression: Maximize Log-likelihood

- Because $L$ is a product, it is convenient to maximize $\ln L$, called the log likelihood
  \[ \ln L(\lambda, x_0) = \sum_{i=1}^{n} x_i \ln x_0 + \lambda t_i x_i - x_0 e^{\lambda t_i} - \ln(x_i!). \]

- Because $x_i$ are constants, drop $\ln(x_i!)$ to maximize
  \[ \ln \tilde{L}(\lambda, x_0) = \sum_{i=1}^{n} x_i \ln x_0 + \lambda t_i x_i - x_0 e^{\lambda t_i}. \]

- This can only be maximized numerically.

- Covariance matrix of the parameters:
  \[ \text{Var}[\lambda, x_0] = \left( D^2 \ln \tilde{L} \right)^{-1} \]
Fitting an Exponential Curve

Confidence Intervals: Likelihood Ratio Test

- To estimate the CI for $\lambda$, we use the likelihood ratio test
- Construct a likelihood profile for $\lambda$
  - $\lambda$ is stepped to both directions of the point estimate $\hat{\lambda}$
  - At each step $k = \pm 1, \pm 2, \cdots$
    - Find $L_k = \max L(x_0 | \lambda_k)$
    - Compute the likelihood ratio
      \[
      D(\lambda_k) = 2 \ln \frac{L_0}{L_k} = 2 \ln L_0 - 2 \ln L_k
      \]
      where $L_0$ is the likelihood at the point estimate.
  - Best practice for step size is to use the standard deviation from the covariance matrix.
  - Approx. $D_k \sim \chi^2_1$, find the 95% CI for $D(\lambda) : (D(\lambda_a), D(\lambda_b))$.
  - The 95% for $\lambda$ is $(\lambda_a, \lambda_b)$.
Negative Binomial Regression

- Poisson regression assumes $E[x_i] = \text{Var}[x_i]$.
- Over-dispersion: $\text{Var}[x_i] > E[x_i]$ because of
  - observation errors
  - non-exponentially distributed waiting times
- Solution: assume that $x_i$ is Negative-Binomial with parameters $r$ and $0 < p < 1$

$$\text{Prob}(x_i | r, p) = \frac{\Gamma(x_i + r)}{x_i! \Gamma(r)} p^r (1 - p)^{x_i}. $$

- Assume the same $r$ for all $x_i$.

$$E[X_i] = \frac{r(1 - p)}{p} \Rightarrow p = \frac{r}{r + E[x_i]} = \frac{r}{r + x_0 e^{\lambda t_i}}$$

- The parameters are $\lambda$, $x_0$, and $r$.
- As $r \to \infty$, the Negative Binomial approaches Poisson.
Application to Simulated Epidemics

The trend of estimated exponential growth rate when using more data points towards the peak of epidemic:

Red: theoretical rate; Black: estimation; blue: 95% CI; grey: epidemic curve
1918 Pandemic Influenza in Philadelphia

The trend of estimated exponential growth rate when using more data points towards the peak of epidemic:

Black: estimation; blue: 95% CI; grey: epidemic curve
Introduction

Exponential Growth Rate Estimate

Some Considerations

Baseline

The early flat phase are non-flu deaths, such deaths are called the baseline P&I deaths.

In a pandemic, most P&I deaths are flu deaths. We can thus ignore the variation in the baseline.

So, we can use a new model for the mean P&I deaths:

\[ x(t) = b + x_0 e^{\lambda t} \]

where \( b \) is the baseline.

We use Poisson regression.

Junling Ma
Department of Mathematics and Statistics, University of Victoria

Estimating the Exponential Growth Rate and \( R_0 \)
1918 Pandemic Influenza in Philadelphia with Baseline

The trend of estimated exponential growth rate when using more data points towards the peak of epidemic:

Black: estimation; blue: 95% CI; grey: epidemic curve
Taking Account of Decreasing Growth Rate

- Exponential growth rate decreases because of the depletion of the susceptibles.
- Use the exponential model,
  - Find the best fitting window by testing goodness of fit.
- Use more sophisticated phenomenological models
  - Logistic model for cumulative cases
  - Richards model for cumulative cases
- Use a mechanistic model, e.g., SIR, SEIR, ...
Single Epidemic Phenomenological Models

- **Logistic model:**
  - The cumulative cases $C(t)$ initially grow exponentially, then approach the final size.
  - The same shape as the logistic model.

$$C'(t) = \lambda C[1 - C/K]$$

This model introduces one more parameter $K$.

- But we should not directly fit the cumulative cases data $c_k \sum_{i=0}^{k} x_k$ to this model, because $c_k$ are not independent.
- Instead, we compute the interval cases $x(t) = c(t + 1) - c(t)$, and fit $x(t)$ to the data $x_i$.

- **Richards model:** cumulative cases has a mean

$$C'(t) = \lambda C[1 - C/K]^\alpha$$
Fit Logistic Model to Simulated Epidemics

Allows the use of more data points:

Red: theoretical rate; black: estimation; blue: 95% confidence interval; grey: epidemic curve
The trend of estimated exponential growth rate when using more data points towards the peak of epidemic:

Black: estimation; blue: 95% confidence interval; grey: epidemic
Process Errors

Same disease parameters may produce different epidemic curves.
Coverage Probability

- Coverage probability of a CI is the probability that CI contains the true parameter value.
- A 95% CI should have 95% coverage probability.
- Because we ignored process errors, this method generally produces narrower confidence intervals.
- Simulations can verify that the coverage probability for incidence cases is poor.
- Larger observation errors, for example, small reporting rates, improve coverage.

Methods that can handle process errors include: one-step ahead, particle filters, MCMC, ...
Estimate $R_0$: SIR Model

First, as an example, we look at an SIR model

$$ S' = -\frac{\beta}{N}SI \ , \ I' = \frac{\beta}{N}SI - \gamma I. $$

► Recall that $\lambda = \beta - \gamma$, so $\beta = \lambda + \gamma$

$$ R_0 = \frac{\beta}{\gamma} = \frac{\lambda + \gamma}{\gamma} = 1 + \frac{\lambda}{\gamma}. $$

► What if $\lambda$ is the exponential growth rate of the incidence curve $X(t) = \beta SI$?
Estimate $R_0$: SEIR Model

\[ S' = -\frac{\beta}{N} Si \, , \, E' = \frac{\beta}{N} Si - \sigma E \, . \, I' = \sigma E - \gamma I \, . \]

- Recall that

\[ \lambda = \rho(J) = \frac{1}{2} \left( \lambda + \gamma + \sqrt{(\sigma - \gamma)^2 + 4\beta\gamma} \right) \]

- Isolate $\beta$,

\[ \beta = \sigma + \frac{\lambda}{\gamma} (\lambda + \gamma + \sigma) \]

\[ R_0 = \frac{\beta}{\gamma} = 1 + \lambda \left( \frac{\lambda}{\sigma \gamma} + \frac{1}{\gamma} + \frac{1}{\sigma} \right) . \]
Estimate $R_0$ with a Model: in General

\[ S' = f(S, I; \theta), \]
\[ I' = g(S, I; \theta). \]

where $S \in \mathbb{R}^m$, $I \in \mathbb{R}^n$, $\theta$ is the vector of parameters.

▶ Recall that the exponential growth rate is the largest eigenvalue of
\[ \frac{\partial g}{\partial I}(S_0, 0; \theta). \]

▶ This relationship usually gives us an estimate of the transmission rate given all the other disease parameters.
▶ $R_0$ can be computed using the inferred transmission rate and all other given parameter values.
Estimate $R_0$ using Generation Interval

See Wallinga and Lipsitch (Proc B 2007, 274:599604)

- Generation interval (serial interval): the waiting time from being infected to secondary infections
  - The generation interval distribution $w(t)$ can be estimated (e.g., from contact tracing) without a mechanistic model.

- Let $n(t)$ is the transmission rate at age of infection $\tau$.
  - $R_0 = \int_0^\infty n(\tau) d\tau$, and $w(\tau) = n(\tau)/R_0$.
  - The incidence curve $x(t) = x(t) \ast n(t)$

- Assume $x(t) = x(0)e^{\lambda t}$,

$$R_0 = \frac{1}{\int_0^\infty e^{-\lambda \tau} w(\tau) d\tau}$$
The Influenza Generation Interval Distribution

The Basic Reproduction Number of 1918 Pandemic Influenza in Philadelphia

Given that we have estimated the exponential growth rate to be

$$\lambda = 0.288, \quad \text{with 95\% confidence interval: (0.286, 0.290)}$$

and with the above generation interval distribution, we can compute that

$$R_0 = 2.16, \quad \text{with 95\% confidence interval: (2.15, 2.17)}$$

This is consistent with other estimations such as Mills et al. (Nature 2004) and Goldstein et al. (PNAS 2009)
Basic Reproduction Number Estimated by Goldstein et al (2009)
Some Considerations

- Deaths v.s. incidences
- Temporal aggregation (e.g., weekly incidences)
- Spatial aggregation (e.g., overall Canada v.s. city level curves)
Some References

- B. M. Bolker (2008), Ecological models and data in R, Princeton University Press
- M. C. J. Bootsma and N. M. Ferguson (2007), PNAS, 104:7588-7593
Some References, Continued

- Y.-H. Hsieh, D. N. Fisman and J. Wu (2010), BMC Research Notes, 3:283
- F. J. Richards (1959), J. Experimental Botany, 10