Least Squares Parameter Estimation

Alun L. Lloyd

Department of Mathematics
Biomathematics Graduate Program
North Carolina State University

Aims of this Lecture

1. Model fitting using least squares

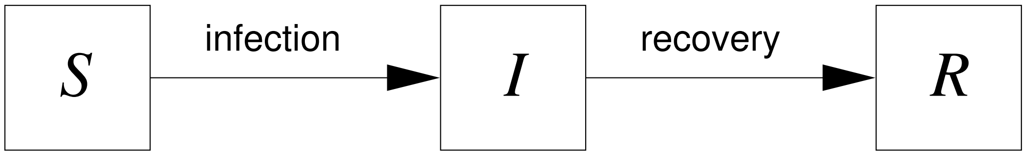
2. Quantification of uncertainty in parameter estimates: asymptotic statistical theory

3. Sensitivity analysis (local linear analysis)

Introduction

SIR model: Mathematical model for the spread of an infection through a population by person-to-person contact

This mathematical model describes the time evolution of the state variables *S*, *I* and *R*, using a set of differential equations, whose terms describe the rates at which various biological processes (infection and recovery) happen



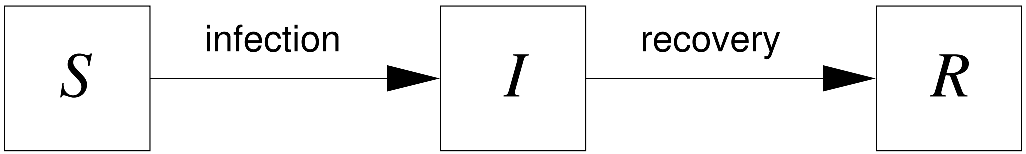
$$dS/dt = -\beta SI/N$$

$$dI/dt = \beta SI/N - \gamma I$$

$$dR/dt = \gamma I.$$

Parameters: constants that appear in the functions describing these rates

The “forward problem”



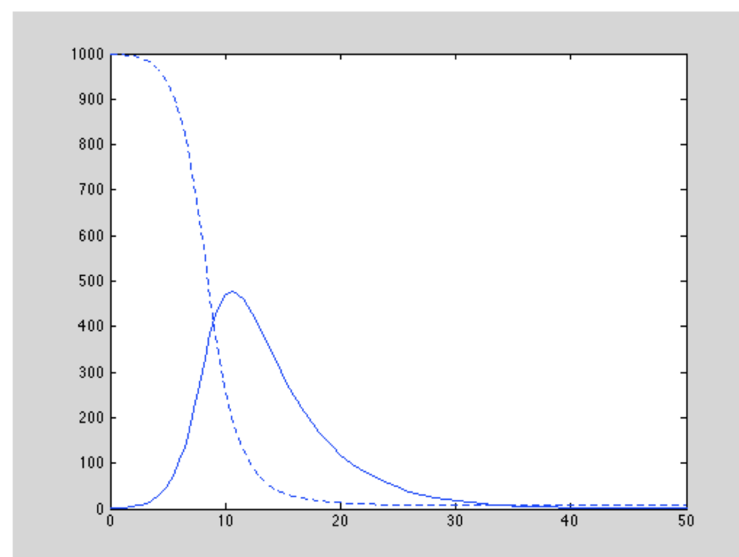
$$dS/dt = -\beta SI/N$$
$$dI/dt = \beta SI/N - \gamma I$$
$$dR/dt = \gamma I.$$

If we know the model equations, the values of the parameters and the initial values of the states, we can (numerically) solve the model forwards in time to find out how the state variables will evolve

Forward simulation of S and I,
given β, γ, N, $S(0)$ and $I(0)$:

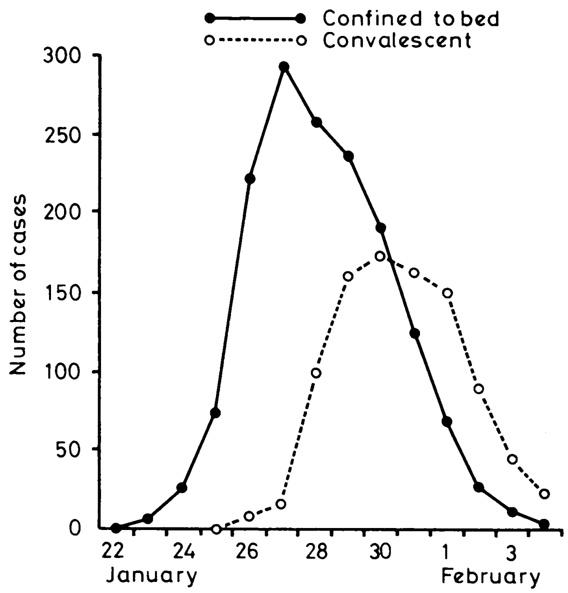
Forward problem:

Parameters $\rightarrow$ dynamical behavior



The "inverse problem" (parameter estimation)

In many cases, we don't know the values of the parameters, but we have observations of the system (e.g. time series data)



Can we figure out parameter values that would make the model exhibit behavior that is consistent with observed data?

Inverse problem: dynamical behavior → parameters

Least-Squares Fitting

One approach to find unknown parameter values: fit a model to time series data

"trajectory matching": find parameter values for which the model provides the best fit to a data set

What do we mean by "best fit"?

Least squares criterion: (ordinary least squares)

Minimize sum of squared differences between model predictions and observed data

- "sum of squared errors"
– hopefully familiar from simple linear regression

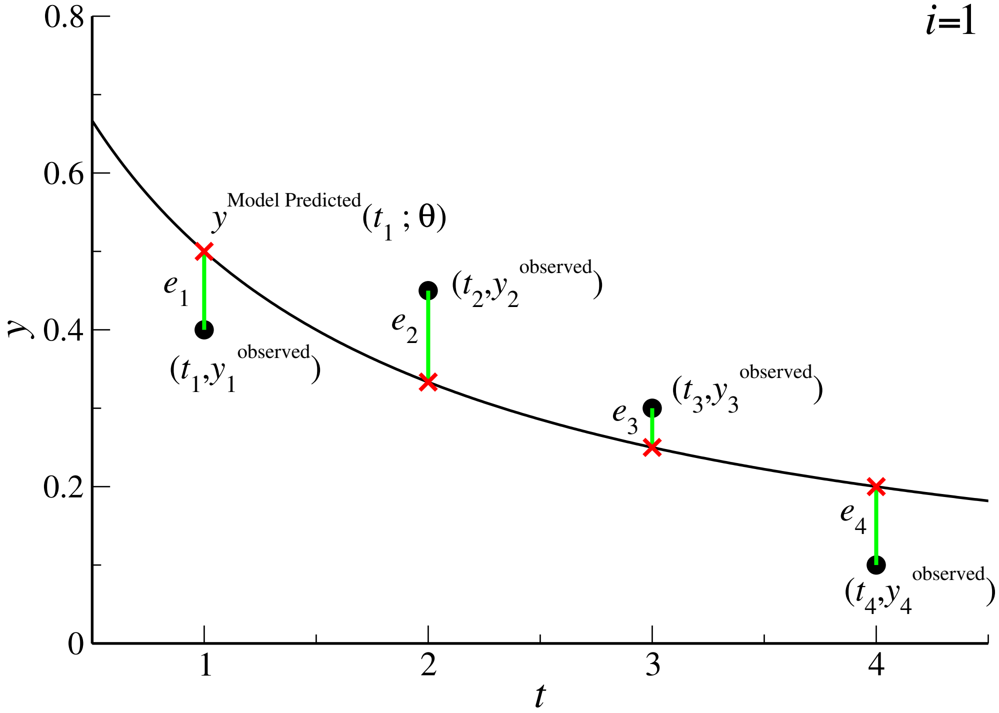
$$F(\theta) = \sum_{i=1}^{n} \left(y^{\text{model predicted}}(t_i;\theta) - y^{\text{observed}}(t_i) \right)^2$$

Here, θ is the vector of parameters

Least-Squares Fitting

Least squares criterion: minimize sum of squared differences between model predictions and observed data (sum of squared errors)

$$F(\theta) = \sum_{i=1}^{n} \left(y^{\text{model predicted}}(t_i;\theta) - y^{\text{observed}}(t_i) \right)^2$$



Dots: data points (observations)

Red crosses: model predictions (for some value of the parameter vector)

Green lines indicate (model-data): called “errors” or “residuals”

[**Why do we use sum of squares?** 1. Cannot simply add errors (cancelation of + and - errors)
2. Sum of squared errors is better behaved than summed absolute errors.]

Least-Squares Fitting

Minimization problem $$\hat{\theta} = \underset{\{\theta \text{ feasible}\}}{\arg\min} \left\{ \sum_{i=1}^{n} \left(y^{\text{model predicted}}(t_i;\theta) - y^{\text{observed}}(t_i) \right)^2 \right\}$$

1. Cannot solve this analytically in general (linear regression is an exception), so will probably need to do optimization numerically
   Process can be non-trivial, e.g. if the function $F(\theta)$ has multiple local minima

2. If our model is an ODE, we probably don't have a formula for the model's output, so will have to numerically solve the ODE to make predictions
   If the initial condition of our ODE is unknown/uncertain, we can include it in the list of quantities to be estimated

3. Estimates depend on the data: would get a different estimate if data was different
   Mathematically: the best-fitting parameter vector is a **random variable**, and its **estimation is a statistical process**

   What are the uncertainties in the estimates we obtain?
   "standard errors" for parameter estimates?

More Generally... State Space Notation

Model: $\dfrac{dx}{dt} = f(x, t; \theta)$, where *x* and *f* are *m* dimensional vectors, and θ is a *p* dimensional vector of parameters

Observation function (maps model to observable quantity):

$$y(t) = h(x, t; \theta)$$

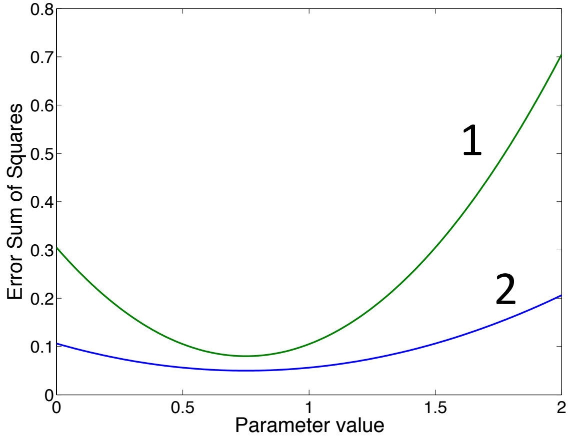
Ordinary least squares parameter estimate is

$$\hat{\theta} = \underset{\{\theta \text{ feasible}\}}{\arg\min} \left\{ \sum_{i=1}^{n} \left(y^{\text{model predicted}}(t_i;\theta) - y^{\text{observed}}(t_i) \right)^2 \right\}$$

Uncertainty Estimates: Intuition

Nonlinear regression theory can be used to provide uncertainty estimates

Intuition: consider two graphs of error sum of squares, based on two different models



Which gives us more certainty in the parameter estimate?

For 2 there is a larger range of parameter values that gives a reasonable fit

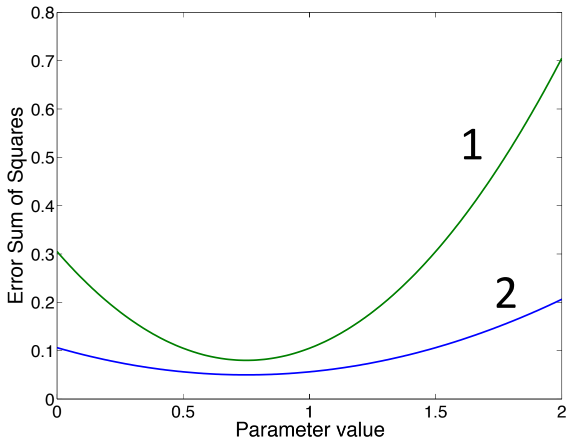
In some sense, there is less information about the true value of the parameter

Intuition: the curvature (2nd derivative) of the SS function at its minimum is an inverse measure of the information

Small second derivative: less information/more uncertainty in value

Idea is formalized in the notion of Fisher information matrix (coming soon)

Uncertainty Estimates: Intuition



A couple of comments:

1. By the chain rule, derivatives of the sum of squares function

$$F(\theta) = \sum_i \left\{ y^{\text{predicted}}(t_i;\theta) - y^{\text{observed}}(t_i) \right\}^2$$

with respect to the parameters will involve derivatives of model predictions with respect to the parameters (we call these sensitivities)

It will be difficult to estimate insensitive parameters

2. Model 2 appears to be a better fit (smaller value for error sum of squares)
   Is there a meaningful notion of how improved the fit is?

Uncertainty Estimates

We need a statistical model for our observations

describes the origin of the "errors" in the data

Simplest statistical model: observations are the values predicted by the model under the "true" parameter values plus observation error

$$y^{\text{observed}}(t_i) = y^{\text{model predicted}}\left(t_i;\theta^{\text{true}}\right) + e_i$$

e_i are the observation errors

In the simplest setting, we assume errors are independent, identically distributed and have constant variance, σ^2

note: following theory is exact if errors are normal, but we don't need to assume errors are normal if we appeal to **large sample size theory** (asymptotic theory)

can also view the following as providing a bound on uncertainty **(see Marisa's talk)**

Uncertainty Estimates

We ask the question:

Suppose we could watch the system evolve over time on multiple occasions, giving us a collection of data sets [because we have different realizations of the noise process], and estimated the parameter values for each of those data sets... How much variation would we see in the parameter estimates?

Could characterize this variation by the **variance** in the values of the estimates

Could also examine the correlation between estimates of different parameters, captured by the **covariance** between their values

All this information is summarized in the covariance (variance/covariance) matrix, Σ

For SIR example: $\Sigma = \begin{pmatrix} \mathrm{Var}(\hat{\beta}) & \mathrm{cov}(\hat{\beta}, \hat{\gamma}) \\ \mathrm{cov}(\hat{\beta}, \hat{\gamma}) & \mathrm{Var}(\hat{\gamma}) \end{pmatrix}$

Uncertainty Estimates for Parameters

Asymptotic (large sample size) theory says that the parameter estimator (a random variable) has a multivariate normal distribution, centered on the true parameter vector and with variance-covariance matrix

$$\Sigma = \sigma^2 \left(\chi(\theta_0)^{\mathrm{T}} \chi(\theta_0)\right)^{-1}$$

Here χ is the *n* x *p* matrix of sensitivities, with entries $\chi(\theta_0)_{i\,j} = \dfrac{\partial y^{\text{predicted}}(t_i;\,\theta_0)}{\partial \theta_j}$

(discuss calculation of sensitivities soon!)

# of data points

# of parameters estimated

$$\chi(\theta_0) = \begin{pmatrix} \frac{\partial y(t_1;\theta_0)}{\partial \theta_1} & \frac{\partial y(t_1;\theta_0)}{\partial \theta_2} & \cdots & \frac{\partial y(t_1;\theta_0)}{\partial \theta_p} \\ \frac{\partial y(t_2;\theta_0)}{\partial \theta_1} & \frac{\partial y(t_2;\theta_0)}{\partial \theta_2} & \cdots & \frac{\partial y(t_2;\theta_0)}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y(t_n;\theta_0)}{\partial \theta_1} & \frac{\partial y(t_n;\theta_0)}{\partial \theta_2} & \cdots & \frac{\partial y(t_n;\theta_0)}{\partial \theta_p} \end{pmatrix}$$

Problem: we don't know θ_0 or σ^2 (true values), so we make use of our estimates

Uncertainty Estimates for Parameters

Asymptotic (large sample size) theory says that the parameter variance-covariance matrix for the estimated parameters is

$$\Sigma = \hat{\sigma}^2 \left(\chi^{(n)}(\hat{\theta})^{\mathrm{T}} \chi^{(n)}(\hat{\theta}) \right)^{-1}$$

Here, $\hat{\theta}$ is our least squares estimate of the parameters

$\hat{\sigma}^2$ is (minimized value of error sum of squares) / (*n* – *p*)

Unbiased estimate

$\chi^{(n)}(\hat{\theta})$ is the *n* x *p* matrix of sensitivities, with entries $\chi^{(n)}(\theta)_{i\,j} = \dfrac{\partial y^{\text{predicted}}(t_i;\,\hat{\theta})}{\partial \theta_j}$

$$\chi^{(n)}(\hat{\theta}) = \begin{pmatrix} \frac{\partial y(t_1;\hat{\theta})}{\partial \theta_1} & \frac{\partial y(t_1;\hat{\theta})}{\partial \theta_2} & \cdots & \frac{\partial y(t_1;\hat{\theta})}{\partial \theta_p} \\ \frac{\partial y(t_2;\hat{\theta})}{\partial \theta_1} & \frac{\partial y(t_2;\hat{\theta})}{\partial \theta_2} & \cdots & \frac{\partial y(t_2;\hat{\theta})}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y(t_n;\hat{\theta})}{\partial \theta_1} & \frac{\partial y(t_n;\hat{\theta})}{\partial \theta_2} & \cdots & \frac{\partial y(t_n;\hat{\theta})}{\partial \theta_p} \end{pmatrix}$$

Uncertainty Estimates for Parameters

Standard errors (SE: standard deviations for the estimated parameters) can be calculated by taking square root of appropriate entry on diagonal entry of Σ

Notice that estimates of different parameters will typically be correlated

Calculate correlation between parameter estimates using $\rho = \dfrac{\mathrm{cov}(\hat{\theta}_i, \hat{\theta}_j)}{\mathrm{SE}(\hat{\theta}_i)\,\mathrm{SE}(\hat{\theta}_j)}$

Sensitivities

(Local) Sensitivity: partial derivative of some quantity of interest with respect to a parameter

In our setting, we will want to calculate sensitivities of the state variables with respect to parameters...

... e.g. if *y*(*t*) = *I*(*t*), we would need to calculate $\frac{\partial I}{\partial \beta}(t)$ and $\frac{\partial I}{\partial \gamma}(t)$

Remember: derivatives relate change in input to change in output:

$$\text{for the derivative } \frac{dy}{dx}, \text{ we have } \Delta y \approx \frac{dy}{dx}\Delta x$$

Sensitivities say how solution curve changes in response to a change in parameter

$$\Delta y(t) \approx \frac{\partial y}{\partial A}(t)\, \Delta A$$

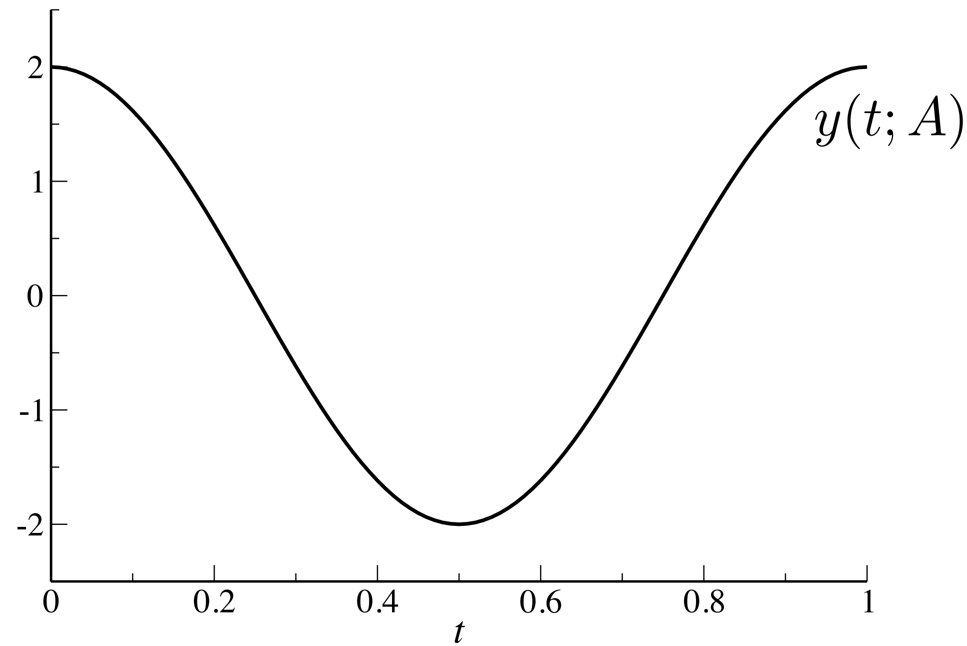
Sensitivities

In our setting, we want to calculate sensitivities of the state variables with respect to parameters:

partial derivatives of state variables with respect to parameters

e.g. $\dfrac{\partial I}{\partial \beta}(t)$ or $\dfrac{dy}{dr}(t)$

Example: $y(t) = A\cos(2\pi t)$



Sensitivities

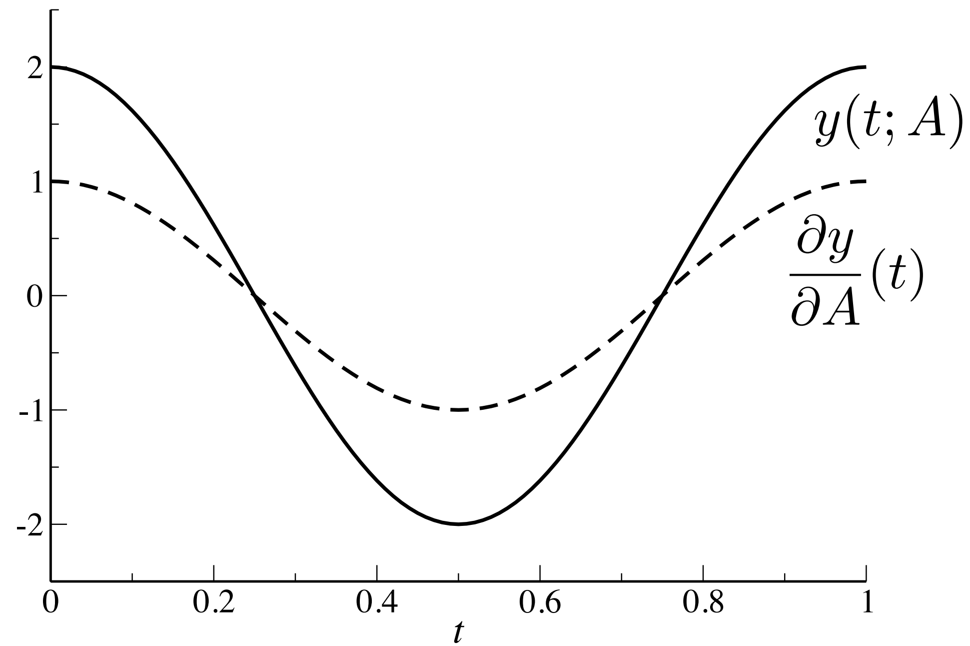
In our setting, we want to calculate sensitivities of the state variables with respect to parameters:

partial derivatives of state variables with respect to parameters

e.g. $\dfrac{\partial I}{\partial \beta}(t)$ or $\dfrac{dy}{dr}(t)$

Example: $y(t) = A\cos(2\pi t)$

Sensitivity: $\dfrac{\partial y}{\partial A}(t) = \cos(2\pi t)$



Sensitivities

In our setting, we want to calculate sensitivities of the state variables with respect to parameters:

partial derivatives of state variables with respect to parameters

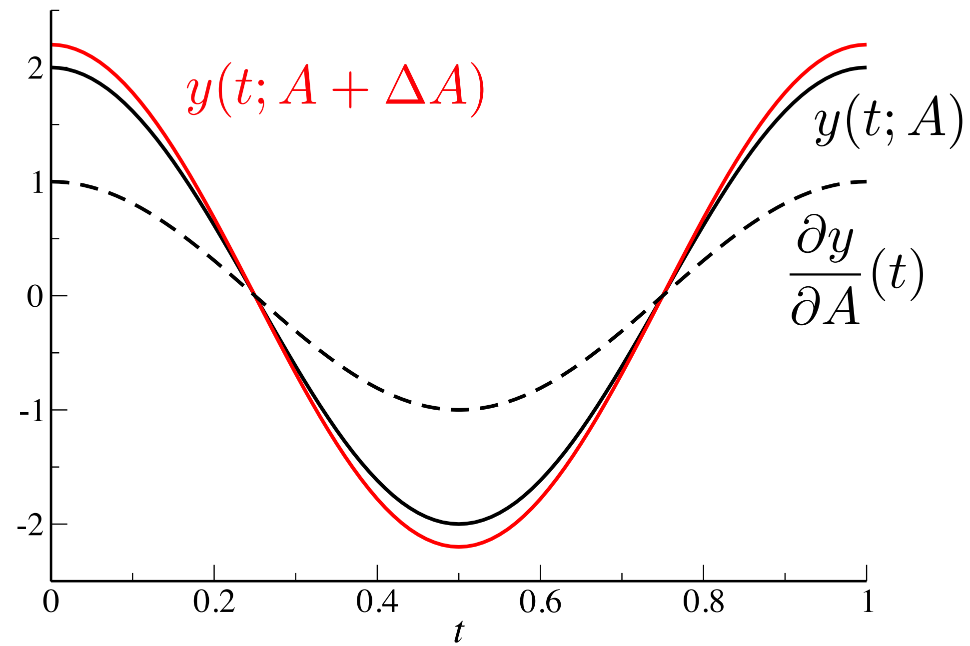
e.g. $\dfrac{\partial I}{\partial \beta}(t)$ or $\dfrac{dy}{dr}(t)$

Example: $y(t) = A\cos(2\pi t)$

Sensitivity: $\dfrac{\partial y}{\partial A}(t) = \cos(2\pi t)$

Interpretation: how does *y*(*t*) change if *A* is increased by a small amount?

$$\Delta y(t) \approx \frac{\partial y}{\partial A}(t)\,\Delta A$$



Sensitivity gives sign and magnitude of change of *y*(*t*) (for any *t*) as *A* is changed

Sensitivities

(Local) Sensitivity: partial derivative of some quantity of interest with respect to a parameter

In our setting, we will want to calculate sensitivities of the state variables with respect to parameters...

... e.g. if $y(t) = I(t)$, we would need to calculate $\frac{\partial I}{\partial \beta}(t)$ and $\frac{\partial I}{\partial \gamma}(t)$

How do we calculate these? If we had a formula for $I(t)$, this would be easy...
... but we don't

Instead we have to use the sensitivity equations

Sensitivity Equations

Sensitivities: partial derivs of state variables with respect to params, e.g. $(\partial I/\partial\gamma)(t)$

For $\dfrac{dx}{dt} = f(x,t;\theta)$ (1) , where *x* and *f* are *m* dimensional (*m* state variables), and θ is a *p* dimensional vector of parameters

We can find the vector of sensitivities $\dfrac{\partial x}{\partial \theta_i}(t)$ of the states with respect to parameter θ_i by differentiating both sides of (1) with respect to the parameter, applying the chain rule, and switching the order of *d/dt* and $d/d\theta_i$ on the left side:

$$\frac{d}{dt}\frac{\partial x}{\partial \theta_i} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta_i} + \frac{\partial f}{\partial \theta_i}$$

(2) : **sensitivity equations**

with initial conditions $\dfrac{\partial x}{\partial \theta_i}(0) = 0_m$

(provided θ_i is a pure parameter, not an initial condition--- see later)

Sensitivity Equations

Sensitivities: partial derivs of state variables with respect to params, e.g. $(\partial I/\partial\gamma)(t)$

For $\dfrac{dx}{dt} = f(x,t;\theta)$ (1) , where *x* and *f* are *m* dimensional (*m* state variables), and θ is a *p* dimensional vector of parameters

We can find the vector of sensitivities $\dfrac{\partial x}{\partial \theta_i}(t)$ of the states with respect to parameter θ_i by differentiating both sides of (1) with respect to the parameter, applying the chain rule, and switching the order of *d*/*dt* and $d/d\theta_i$ on the left side:

$$\frac{d}{dt}\frac{\partial x}{\partial \theta_i} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta_i} + \frac{\partial f}{\partial \theta_i}$$

(2) : **sensitivity equations**

with initial conditions $\dfrac{\partial x}{\partial \theta_i}(0) = 0_m$

(provided θ_i is a pure parameter, not an initial conditio

Why do we need the chain rule? Because solution *x*(*t*) depends on parameter θ_i

Sensitivity Equations

Sensitivities: partial derivs of state variables with respect to params, e.g. $(\partial I/\partial\gamma)(t)$

For $\dfrac{dx}{dt} = f(x, t; \theta)$ (1) , where *x* and *f* are *m* dimensional (*m* state variables), and θ is a *p* dimensional vector of parameters

We can find the vector of sensitivities $\dfrac{\partial x}{\partial \theta_i}(t)$ of the states with respect to parameter θ_i by differentiating both sides of (1) with respect to the parameter, applying the chain rule, and switching the order of *d/dt* and $d/d\theta_i$ on the left side:

$$\frac{d}{dt}\frac{\partial x}{\partial \theta_i} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta_i} + \frac{\partial f}{\partial \theta_i}$$

(2) : **sensitivity equations**

with initial conditions $\dfrac{\partial x}{\partial \theta_i}(0) = 0_m$

Why are these zero? Initial conditions don't depend on parameter value

(provided θ_i is a pure parameter, not an initial conditio

Sensitivity Equations

Sensitivities: partial derivs of state variables with respect to params, e.g. $(\partial I/\partial\gamma)(t)$

For $\dfrac{dx}{dt} = f(x,t;\theta)$ (1) , where x and f are m dimensional (m state variables), and θ is a p dimensional vector of parameters

We can find the vector of sensitivities $\dfrac{\partial x}{\partial \theta_i}(t)$ of the states with respect to parameter θ_i by differentiating both sides of (1) with respect to the parameter, applying the chain rule, and switching the order of d/dt and $d/d\theta_i$ on the left side:

$$\frac{d}{dt}\frac{\partial x}{\partial \theta_i} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta_i} + \frac{\partial f}{\partial \theta_i}$$ (2) :

with initial conditions $\dfrac{\partial x}{\partial \theta_i}(0) = 0_m$

(provided θ_i is a pure parameter, not an i

The matrix $\dfrac{\partial f}{\partial x}$ is the Jacobian matrix

- differentiate RHS of ODE w.r.t. state vars.

$\dfrac{\partial f}{\partial \theta_i}$ is derivative of RHS w.r.t. parameter

This is a generalization of linear stability analysis

Sensitivity Equations

Sensitivities: partial derivs of state variables with respect to params, e.g. $(\partial I/\partial\gamma)(t)$

For $\dfrac{dx}{dt} = f(x, t; \theta)$ (1) , where *x* and *f* are *m* dimensional (*m* state variables), and θ is a *p* dimensional vector of parameters

The sensitivity equations for $\dfrac{\partial x}{\partial \theta_i}(t)$ are $\dfrac{d}{dt}\dfrac{\partial x}{\partial \theta_i} = \dfrac{\partial f}{\partial x}\dfrac{\partial x}{\partial \theta_i} + \dfrac{\partial f}{\partial \theta_i}$

with initial conditions $\dfrac{\partial x}{\partial \theta_i}(0) = 0_m$

1. Sensitivity eqns for different states with respect to a given parameter are coupled

2. Sensitivity eqns across different parameters are not coupled...

... so can arrange sensitivities with respect to different parameters into a matrix

Sensitivity Equations

For $\dfrac{dx}{dt} = f(x,t;\theta)$ (1) , where *x* and *f* are *m* dimensional (*m* state variables), and θ is a *p* dimensional vector of parameters

the *m* by *p* matrix of sensitivities, $\dfrac{\partial x}{\partial \theta}(t)$, satisfies the ODE system

$$\frac{d}{dt}\frac{\partial x}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial \theta} \qquad (2)$$

with initial conditions $\dfrac{\partial x}{\partial \theta}(0) = 0_{m \times p}$

Equations (2) are the sensitivity equations for the system

Sensitivity matrix: Sensitivities of different states with respect to a given parameter are arranged in a single column, different columns depict sensitivities with respect to different parameters

Sensitivity Equations

For $\dfrac{dx}{dt} = f(x, t; \theta)$ (1) , where *x* and *f* are *m* dimensional (*m* state variables), and θ is a *p* dimensional vector of parameters

the *m* by *p* matrix of sensitivities, $\dfrac{\partial x}{\partial \theta}(t)$, satisfies the ODE system

$$\frac{d}{dt}\frac{\partial x}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial \theta} \qquad (2)$$

with initial conditions $\dfrac{\partial x}{\partial \theta}(0) = 0_{m \times p}$

Equations (2) are the sensitivity equations for the system

The Jacobian matrix $\dfrac{\partial f}{\partial x}$ depends on the state variables, so the sensitivity equations (2)

[a linear system of ODEs] must be solved simultaneously with the governing equations (1)

Sensitivities with Respect to Initial Conditions?

If an initial condition is unknown, we could include it as one of the quantities to be estimated using least squares

Everything goes as before, except that sensitivity equations are a little different if you are looking at sensitivities with respect to an initial condition

$$\frac{d}{dt}\frac{\partial x}{\partial \theta_j} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta_j}$$

We don't have that second term on the right side because the equations right hand sides of our model don't involve initial conditions

$$\frac{\partial x}{\partial \theta_j}(0) = \mathbf{e}_j$$

Unit vector with 1 in *j*'th place, 0 elsewhere
Derivative of initial condition *i* with respect to initial condition *i* is one, zero with respect to other initial conditions

So, we now know how to calculate sensitivities with respect to either parameters or initial conditions; we can calculate both together

Sensitivity Equations for Logistic Growth Model

Because we have an analytic solution of the logistic growth model

$$\frac{dy}{dt} = ry\left(1 - \frac{y}{K}\right); \qquad y(0) = y_0$$

$$y(t) = \frac{K}{1 + (K/y_0 - 1)\, e^{-rt}}$$

We can calculate the sensitivities of y(t) with respect to K , r and y_0 using calculus

But we could also calculate those three sensitivities using the sensitivity equations method... and check that the two methods give the same answer!

Next slide outlines one of those calculations, for $\partial y/\partial r$

Sensitivity Equations for Logistic Growth Model

Let's set up sensitivity equation for $\partial y/\partial r$, translating from general notation

General Notation

$$\frac{dx}{dt} = f(x,t;\theta)$$

Notation of our problem

$$\frac{dy}{dt} = ry\left(1 - \frac{y}{K}\right)$$

"*x*" is "*y*", and "θ" is "*r*", "*f*" is *ry*(1-*y*/*K*)

$$\frac{d}{dt}\frac{\partial x}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial \theta}$$

$$\frac{d}{dt}\frac{\partial y}{\partial r} = \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} + \frac{\partial f}{\partial r}$$

Calculating ∂*f*/∂*y* and ∂*f*/∂*r*, and substituting:

$$\frac{d}{dt}\frac{\partial y}{\partial r} = r\left(1 - \frac{2y}{K}\right)\frac{\partial y}{\partial r} + y\left(1 - \frac{y}{K}\right)$$

Have to integrate a two dimensional system, with states *y* and $\partial y/\partial r$, with initial value of sensitivity = 0

Derivation of the Sensitivity Equations for the SIR Model
(Appendix of Capaldi et al. 2012)

We have

$$dS/dt = -\beta SI/N$$

$$dI/dt = \beta SI/N - \gamma I$$

The four sensitivities, arranged in a matrix:

$$\frac{\partial x}{\partial \theta} = \begin{pmatrix} \frac{\partial S}{\partial \beta} & \frac{\partial S}{\partial \gamma} \\ \frac{\partial I}{\partial \beta} & \frac{\partial I}{\partial \gamma} \end{pmatrix}$$

Jacobian:

$$\frac{\partial f}{\partial x} = J(S,I) = \begin{pmatrix} \frac{\partial}{\partial S}(dS/dt) & \frac{\partial}{\partial I}(dS/dt) \\ \frac{\partial}{\partial S}(dI/dt) & \frac{\partial}{\partial I}(dI/dt) \end{pmatrix} = \begin{pmatrix} -\beta I/N & -\beta S/N \\ \beta I/N & \beta S/N - \gamma \end{pmatrix}$$

Derivative of right hand sides of differential equations with respect to parameters

$$\frac{\partial f}{\partial \theta} = \begin{pmatrix} \frac{\partial}{\partial \beta}\left(\frac{dS}{dt}\right) & \frac{\partial}{\partial \gamma}\left(\frac{dS}{dt}\right) \\ \frac{\partial}{\partial \beta}\left(\frac{dI}{dt}\right) & \frac{\partial}{\partial \gamma}\left(\frac{dI}{dt}\right) \end{pmatrix} = \begin{pmatrix} -SI/N & 0 \\ SI/N & -I \end{pmatrix}$$

Derivation of the Sensitivity Equations for the SIR Model

(Appendix of Capaldi et al. 2012)

We have $dS/dt = -\beta SI/N$

$dI/dt = \beta SI/N - \gamma I$

Shorthand notation for sensitivities

The four sensitivities, arranged in a matrix: $\dfrac{\partial x}{\partial \theta} = \begin{pmatrix} \dfrac{\partial S}{\partial \beta} & \dfrac{\partial S}{\partial \gamma} \\ \dfrac{\partial I}{\partial \beta} & \dfrac{\partial I}{\partial \gamma} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ \phi_3 & \phi_4 \end{pmatrix}$

Jacobian: $\dfrac{\partial f}{\partial x} = J(S,I) = \begin{pmatrix} \frac{\partial}{\partial S}(dS/dt) & \frac{\partial}{\partial I}(dS/dt) \\ \frac{\partial}{\partial S}(dI/dt) & \frac{\partial}{\partial I}(dI/dt) \end{pmatrix} = \begin{pmatrix} -\beta I/N & -\beta S/N \\ \beta I/N & \beta S/N - \gamma \end{pmatrix}$

Derivative of right hand sides of differential equations with respect to parameters $\dfrac{\partial f}{\partial \theta} = \begin{pmatrix} \frac{\partial}{\partial \beta}\left(\frac{dS}{dt}\right) & \frac{\partial}{\partial \gamma}\left(\frac{dS}{dt}\right) \\ \frac{\partial}{\partial \beta}\left(\frac{dI}{dt}\right) & \frac{\partial}{\partial \gamma}\left(\frac{dI}{dt}\right) \end{pmatrix} = \begin{pmatrix} -SI/N & 0 \\ SI/N & -I \end{pmatrix}$

Derivation of the Sensitivity Equations for the SIR Model
(Appendix of Capaldi et al. 2012)

Putting this all into sensitivity equations (2):

$$\frac{d}{dt}\begin{pmatrix} \phi_1 & \phi_2 \\ \phi_3 & \phi_4 \end{pmatrix} = \begin{pmatrix} -\beta I/N & -\beta S/N \\ \beta I/N & \beta S/N - \gamma \end{pmatrix}\begin{pmatrix} \phi_1 & \phi_2 \\ \phi_3 & \phi_4 \end{pmatrix} + \begin{pmatrix} -SI/N & 0 \\ SI/N & -I \end{pmatrix}$$

Four sensitivity equations, arranged in matrix form

Derivation of the Sensitivity Equations for the SIR Model

(Appendix of Capaldi et al. 2012)

Putting this all into sensitivity equations (2):

$$\frac{d}{dt}\begin{pmatrix} \boxed{\phi_1} & \phi_2 \\ \phi_3 & \phi_4 \end{pmatrix} = \begin{pmatrix} \boxed{-\beta I/N \quad -\beta S/N} \\ \beta I/N \quad \beta S/N - \gamma \end{pmatrix} \begin{pmatrix} \boxed{\begin{matrix}\phi_1 \\ \phi_3\end{matrix}} & \begin{matrix}\phi_2 \\ \phi_4\end{matrix} \end{pmatrix} + \begin{pmatrix} \boxed{-SI/N} & 0 \\ SI/N & -I \end{pmatrix}$$

Four sensitivity equations, arranged in matrix form

Notice for this 2D system, the equations for $\partial S/\partial\beta$ (aka ϕ_1) and $\partial I/\partial\beta$ (aka ϕ_3) are coupled, so must be solved together [similarly for the two sensitivities wrt γ]

Derivation of the Sensitivity Equations for the SIR Model

(Appendix of Capaldi et al. 2012)

Putting this all into sensitivity equations (2):

$$\frac{d}{dt}\begin{pmatrix} \phi_1 & \phi_2 \\ \phi_3 & \phi_4 \end{pmatrix} = \begin{pmatrix} -\beta I/N & -\beta S/N \\ \beta I/N & \beta S/N - \gamma \end{pmatrix}\begin{pmatrix} \phi_1 & \phi_2 \\ \phi_3 & \phi_4 \end{pmatrix} + \begin{pmatrix} -SI/N & 0 \\ SI/N & -I \end{pmatrix}$$

Four sensitivity equations, arranged in matrix form

Notice for this 2D system, the equations for $\partial S/\partial\beta$ (aka ϕ_1) and $\partial I/\partial\beta$ (aka ϕ_3) are coupled, so must be solved together

As mentioned previously, these have to be solved simultaneously with the original model equations

$$dS/dt = -\beta SI/N$$

$$dI/dt = \beta SI/N - \gamma I$$

Initial conditions for the sensitivities: all four equal zero at $t = 0$

[at t=0, all state variables equal their initial values. No dependence of this on the parameters, so those four partial derivatives are zero.]

Behavior of the Sensitivity Equations

Capaldi et al. 2012: plotted curves of $\partial I/\partial\beta$ (solid) and $\partial I/\partial\gamma$ (dashed)

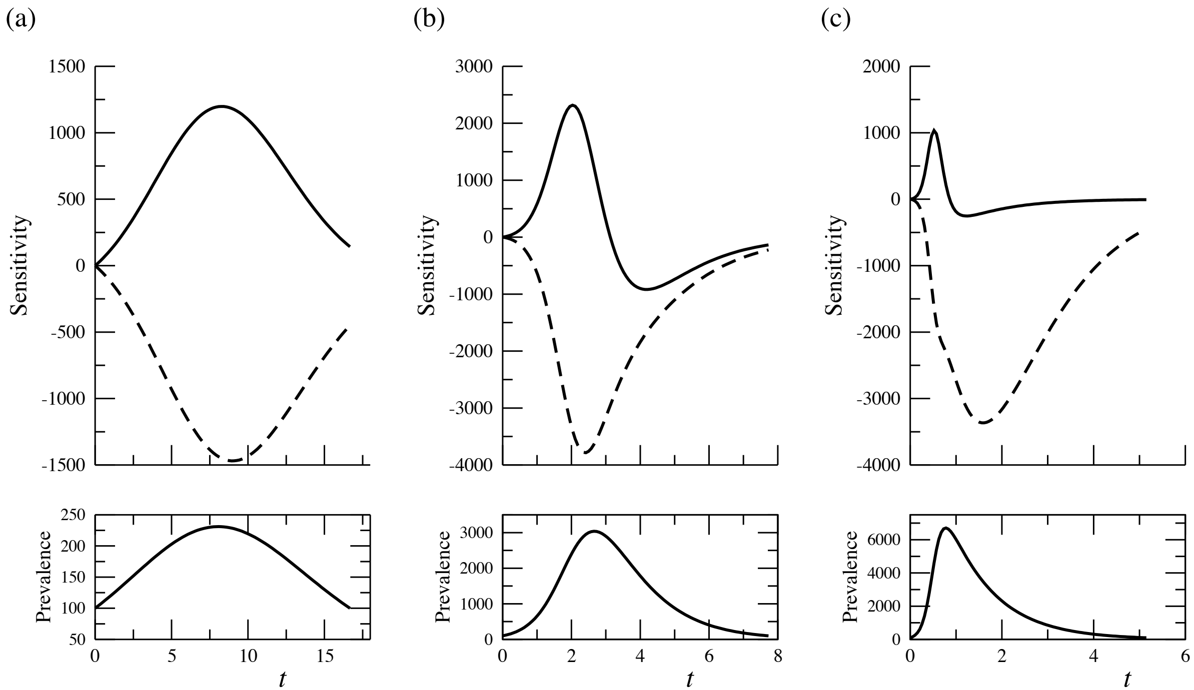
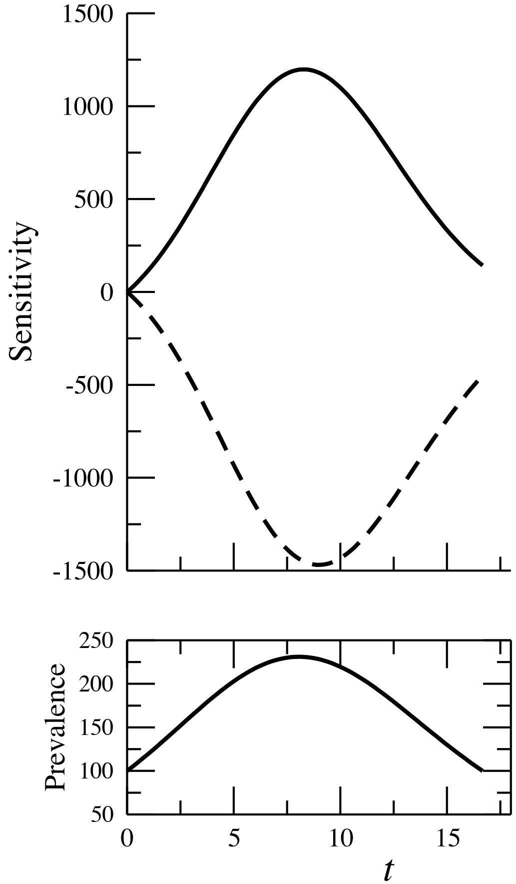


FIGURE 3. Sensitivities of $I(t)$ (*i.e.*, prevalence) with respect to the model parameters β (solid curves) and γ (dashed curves) are shown on the upper panels of the graphs for a) $R_0 = 1.2$, b) $R_0 = 3$ and c) $R_0 = 10$. The lower panel of each graph displays the corresponding prevalence-time curve. The initial conditions of the SIR model were $S_0 = 9900$, $I_0 = 100$, with $N = 10,000$ and γ was taken equal to one, so $\beta = R_0$.

Non-Identifiability of Parameters

Capaldi et al. 2012 curves of $\partial I/\partial\beta$ (solid) and $\partial I/\partial\gamma$ (dashed) for R_0=1.2

(a)



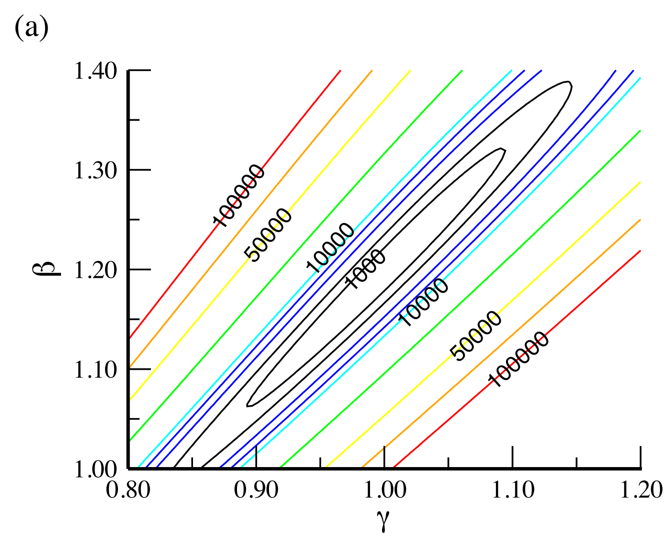
The two sensitivity curves are almost mirror images of each other…

Increasing β by a small amount has more or less the exact opposite effect on *I*(*t*) as does increasing γ by that small amount…

…so we can get an almost identical model fit if we increase β and γ by the same amount, making it difficult to estimate the two separately

Correlations between parameter estimates

If we were to look at the sum of squares function for this situation:



Notice elliptical-shaped contours near the minimum point:

SS varies rapidly parallel to the minor axis

SS varies slowly parallel to the major axis,

It's easy to locate the minimum point in one direction, difficult in the other

> In this case, we can easily estimate the ratio of β/γ, but it's more difficult to estimate their separate values

This is linked to the notion of **identifiability** : whether one can separately estimate different parameters of the model

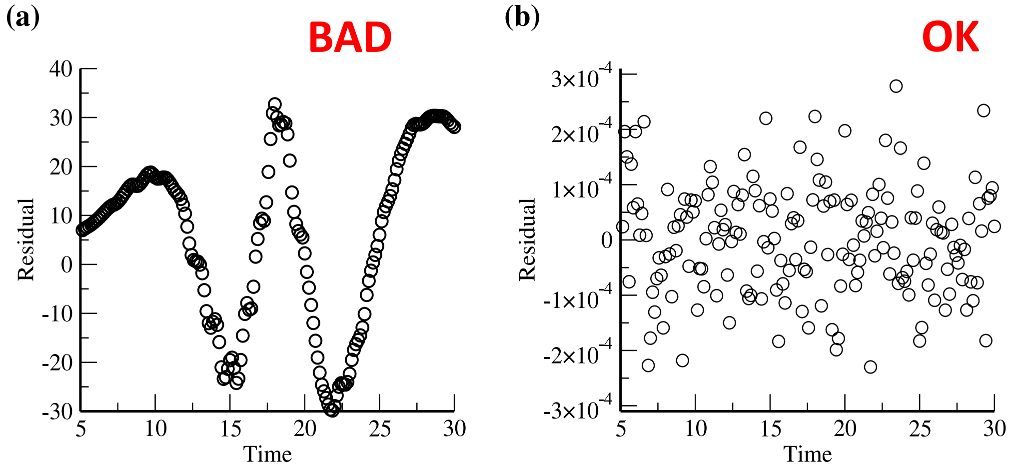
Diagnostic Plots

Statistical model assumed that errors (residuals) were independent and identically distributed

Check this by plotting residuals, both against time and against predicted value

You do not want to see clear patterns in these plots

e.g. two plots of residuals vs time from Cintron-Arias et al. (2009):



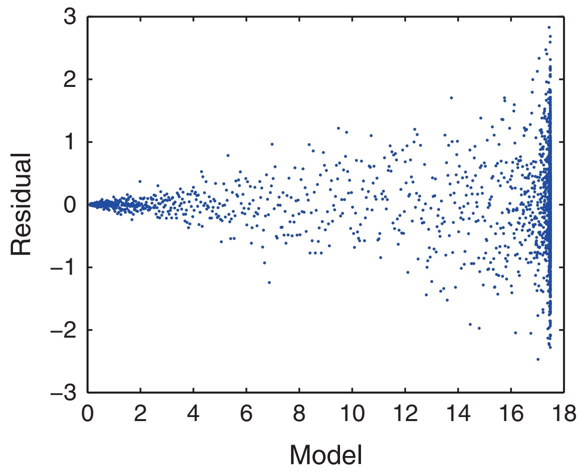
Clear temporal pattern in residuals/time plot in (a) suggests systematic problem with the model fit, e.g. model mis-specification

Diagnostic Plots

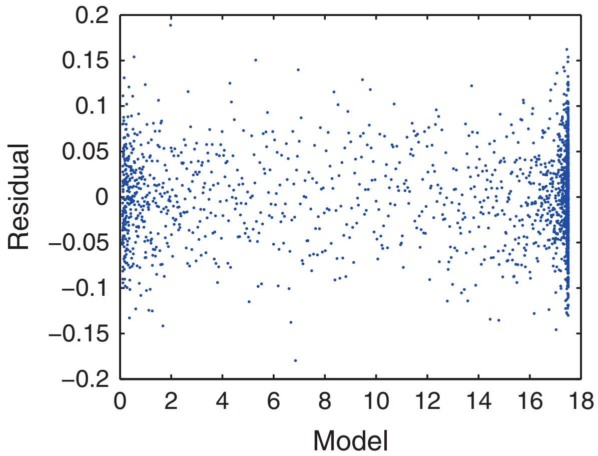
Plots of residuals vs model prediction should not show a pattern

e.g. from Banks et al. 2009

BAD



OK



In first panel, magnitude of observation errors clearly increases with model prediction (or size of observation), violating the assumptions of the statistical model

In such cases, we should use **weighted least squares** (see, e.g., Banks et al. 2009)

Comparing Fits of Different Models

Could fit an SEIR model to our outbreak data

involves one additional parameter

Which would fit better: SIR or SEIR model?

SIR model is special case ($\sigma \to \infty$) of the SEIR model, so SEIR model is guaranteed to be able to fit better

In general, more flexible models (more parameters) are able to better fit a given dataset… but can be prone to overfitting data

[e.g. if you have *n* points on an x-y plot, can find an *n*-1 degree polynomial that goes **exactly** through all points]

“With four parameters I can fit an elephant, and with five I can make him wiggle his trunk” (attributed to John von Neumann)

Information criteria (e.g. Akaike Information Criterion, AIC) balance goodness of fit vs model complexity, penalizing models with more fitted parameters

Should have many more data points than parameters to be fitted!

Limitations of the least-squares approach

Important point: the statistical model used here assumes that observation noise is the only source of error

In reality:

1. The disease process is stochastic. This uncertainty will propagate through the process, potentially increasing over time

   There are approaches that can account for “process noise”
   (e.g particle filter)

2. Model mis-specification. Unlikely that we have exactly the right model.

   ???

Diagnostic plots would likely indicate issues…

A Final Point

Just because a model provides a "good" fit to data does not mean that the model correctly characterizes the processes governing a system

a model can fit, but for the wrong reasons

References

Anonymous (1978). Influenza in a boarding school. Brit. Med. J. **6112**, 587.

Banks, H.T., et al. (2009). An inverse problem statistical methodology summary. In Chowell et al., eds. Mathematical and statistical estimation approaches in epidemiology, Springer.

Banks, H.T., et al. (2014). Modeling and Inverse Problems in the Presence of Uncertainty. CRC Press.

Capaldi, A., et al. (2012). Parameter estimation and uncertainty quantification for an epidemic model. Math. Biosci. Eng. **9**, 553-576.

Cintron-Arias, A., et al. (2009). The estimation of the effective reproductive number from disease outbreak data. Math. Biosci. Eng. 6, 261-282.

Lloyd, A.L. (2001). The dependence of viral parameter estimates on the assumed viral life cycle: limitations of studies of viral load data. Proc. R. Soc. Lond. B. **268**, 847-854.

Nowak, M.A., et al. (1997). Viral dynamics of primary viremia and antiretroviral therapy in simian immunodeficiency virus infection. J. Virol. 71, 7518-7525.