A Practical Algorithm for "Practical" Parameter Identifiability Analysis



Yasunori Aoki

University of Waterloo

Ben Holder

Ryerson University

Hans De Sterck

University of Waterloo

Ken Hayami National Institute of Informatics

A Practical Algorithm for "Practical" Parameter Identifiability Analysis



Yasunori Aoki

University of Waterloo

Ben Holder

Ryerson University

Hans De Sterck

University of Waterloo

Ken Hayami National Institute of Informatics

Parameter Identifiability Analysis

Parameter Identifiability Analysis

Assuming there is no variability of data, can we uniquely estimate model parameters?

Assuming there is no variability of data, can we uniquely estimate model parameters?

Practical Parameter Identifiability Analysis

Assuming there is no variability of data, can we uniquely estimate model parameters?

Practical Parameter Identifiability Analysis

Given a certain level of variability of data, how precisely can we estimate model parameters?

Nonlinear Model	Computational Cost

	Nonlinear Model	Computational Cost
Fisher Information Matrix		

	Nonlinear Model	Computational Cost
Fisher Information Matrix	not-Robust	



	Nonlinear Model	Computational Cost
Fisher Information Matrix	not-Robust	Fast
Monte Carlo / Bootstrap method		







$$\frac{du_1}{dt} = -x_1 u_1 u_4$$
$$\frac{du_2}{dt} = x_1 u_1 u_4 - \frac{1}{x_2} u_2$$
$$\frac{du_3}{dt} = \frac{1}{x_2} u_2 - \frac{1}{x_3} u_3$$
$$\frac{du_4}{dt} = \frac{x_4}{x_5} u_3 - x_6 u_4$$

$$u_1(t = 0) = x_5$$

 $u_2(t = 0) = 0$
 $u_3(t = 0) = 0$
 $u_4(t = 0) = x_7$

$$\frac{du_1}{dt} = -x_1 u_1 u_4$$
$$\frac{du_2}{dt} = x_1 u_1 u_4 - \frac{1}{x_2} u_2$$
$$\frac{du_3}{dt} = \frac{1}{x_2} u_2 - \frac{1}{x_3} u_3$$
$$\frac{du_4}{dt} = \frac{x_4}{x_5} u_3 - x_6 u_4$$
$$u_1(t=0) = x_5$$
$$u_2(t=0) = 0$$
$$u_3(t=0) = 0$$

 $u_4(t=0) = x_7$

$$\frac{du_1}{dt} = -x_1 u_1 u_4$$
$$\frac{du_2}{dt} = x_1 u_1 u_4 - \frac{1}{x_2} u_2$$
$$\frac{du_3}{dt} = \frac{1}{x_2} u_2 - \frac{1}{x_3} u_3$$
$$\frac{du_4}{dt} = \frac{x_4}{x_5} u_3 - x_6 u_4$$
$$u_1(t=0) = x_5$$
$$u_2(t=0) = 0$$
$$u_3(t=0) = 0$$
$$u_4(t=0) = x_7$$









Find \boldsymbol{x} that minimizes $||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}^*_{\mathrm{exp1}}||_2$



Find \boldsymbol{x} that minimizes $||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}^*_{\exp 2}||_2$



Find \boldsymbol{x} that minimizes $||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}^*_{\exp 3}||_2$

We do not usually have enough sets of experimental data to conduct statistical analyses on the sets of parameters found from them...

We do not usually have enough sets of experimental data to conduct statistical analyses on the sets of parameters found from them...

Artificially create data.

We do not usually have enough sets of experimental data to conduct statistical analyses on the sets of parameters found from them...

Artificially create data.

Explicitly specify variability level and distribution

Monte Carlo method

We do not usually have enough sets of experimental data to conduct statistical analyses on the sets of parameters found from them...

Artificially create data.

Explicitly specify variability level and distribution

Resample residuals

Monte Carlo method

Bootstrap method



Find \boldsymbol{x} that minimizes $||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}^*_{\mathrm{exp1}}||_2$

Find x that minimizes $||f(x) - y^*_{exp1}||_2$ Find x that minimizes $||f(x) - y^*_{exp2}||_2$
Find x that minimizes $||f(x) - y^*_{exp1}||_2$ Find x that minimizes $||f(x) - y^*_{exp2}||_2$ Find x that minimizes $||f(x) - y^*_{exp3}||_2$

٠

•

Find x that minimizes $||f(x) - y^*_{exp1}||_2$ Find x that minimizes $||f(x) - y^*_{exp2}||_2$ Find x that minimizes $||f(x) - y^*_{exp3}||_2$

Find \boldsymbol{x} that minimizes $||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}^*_{\mathrm{exp1000}}||_2$

Find \boldsymbol{x} that minimizes $||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}^*||_2$

Find \boldsymbol{x} that minimizes $||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}^*||_2$

Gauss Newton method $\boldsymbol{x}_{\text{new}} = \boldsymbol{x}_{\text{old}} + (J^{\text{T}}J)^{-1}(J^{\text{T}}(\boldsymbol{f}(\boldsymbol{x}_{\text{old}}) - \boldsymbol{y}^{*}))$

Find \boldsymbol{x} that minimizes $||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}^*||_2$

Levenberg-Marquardt method $\boldsymbol{x}_{\text{new}} = \boldsymbol{x}_{\text{old}} + (J^{\text{T}}J + \lambda I)^{-1}(J^{\text{T}}(\boldsymbol{f}(\boldsymbol{x}_{\text{old}}) - \boldsymbol{y}^{*}))$

Find \boldsymbol{x} that minimizes $||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}^*||_2$

Levenberg-Marquardt method $\boldsymbol{x}_{\text{new}} = \boldsymbol{x}_{\text{old}} + (J^{\text{T}}\boldsymbol{J} + \lambda I)^{-1}(J^{\text{T}}(\boldsymbol{f}(\boldsymbol{x}_{\text{old}}) - \boldsymbol{y}^{*}))$

Find \boldsymbol{x} that minimizes $||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}^*||_2$

Levenberg-Marquardt method $\boldsymbol{x}_{\text{new}} = \boldsymbol{x}_{\text{old}} + (J^{\text{T}}\boldsymbol{J} + \lambda I)^{-1}(J^{\text{T}}(\boldsymbol{f}(\boldsymbol{x}_{\text{old}}) - \boldsymbol{y}^{*}))$

Approximating the Jacobian matrix is computationally **expensive**...









Jacobian matrix:



 $\begin{bmatrix} \frac{\partial f_1}{\partial x_1} \\ \frac{\partial f_2}{\partial x_1} \\ \vdots \\ \frac{\partial f_n}{\partial x_1} \end{bmatrix} \approx f(x_1, x_2, \cdots x_m)$



$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} \\ \frac{\partial f_2}{\partial x_1} \\ \vdots \\ \frac{\partial f_n}{\partial x_1} \end{bmatrix} \approx f(x_1, x_2, \cdots x_m) - f(x_1 - \epsilon, x_2, \cdots x_m)$$



 $\begin{bmatrix} \frac{\partial f_1}{\partial x_1} \\ \frac{\partial f_2}{\partial x_1} \\ \vdots \\ \frac{\partial f_n}{\partial x_n} \end{bmatrix} \approx \frac{f(x_1, x_2, \cdots x_m) - f(x_1 - \epsilon, x_2, \cdots x_m)}{\epsilon}$





Number of simulations needed to obtain 1,000 sets of parameters.

Number of simulations needed to obtain 1,000 sets of parameters.

(Number of parameters + 1)

Number of simulations needed to obtain 1,000 sets of parameters.

(Number of parameters + 1) x Number of iterations

Number of simulations needed to obtain 1,000 sets of parameters.

(Number of parameters + 1) x Number of iterations x 1,000







Approximating the Jacobian matrix using the **neighbours**...

Approximating the Jacobian matrix using the **neighbours**...

Approximating the Jacobian matrix using the **neighbours**...

Jacobian is like...

Slope

Approximating the Jacobian matrix using the **neighbours**...

Jacobian is like...



Approximating the Jacobian matrix using the **neighbours**...

Jacobian is like... Slope $\approx \frac{\text{Rise}}{\text{Run}}$

Rise:
$$\Delta y = f(x) - f(x_{\text{neighbour}})$$

Approximating the Jacobian matrix using the **neighbours**...

Jacobian is like... Slope $\approx ---$ Run

Rise:
$$\Delta y = f(x) - f(x_{\text{neighbour}})$$

Run: $\Delta x = x - x_{\text{neighbour}}$

Approximating the Jacobian matrix using the **neighbours**...

Jacobian is like... $J_{approx} = rac{\Delta y}{\Delta x}$

Rise:
$$\Delta y = f(x) - f(x_{\text{neighbour}})$$

Run: $\Delta x = x - x_{\text{neighbour}}$

Approximating the Jacobian matrix using the **neighbours**...

$$\Delta \boldsymbol{y} = J_{\mathrm{approx}} \Delta \boldsymbol{x}$$

$$\Delta \boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_{\text{neighbour}})$$

$$\Delta \boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_{\text{neighbour}}$$

Approximating the Jacobian matrix using the **neighbours**...

$$\Delta \boldsymbol{y}^{(i)} = J_{\mathrm{approx}} \Delta \boldsymbol{x}^{(i)}$$

for $i = 1, 2, \ldots, n_{\text{neighbour}}$

$$\Delta \boldsymbol{y}^{(i)} = \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_{ ext{neighbour}}^{(i)})$$

 $\Delta \boldsymbol{x}^{(i)} = \boldsymbol{x} - \boldsymbol{x}_{ ext{neighbour}}^{(i)}$






















Number of simulations needed to obtain 1,000 sets of parameters.

Number of simulations needed to obtain 1,000 sets of parameters.

1

Number of simulations needed to obtain 1,000 sets of parameters.

1 x Number of iterations

Number of simulations needed to obtain 1,000 sets of parameters.

1 x Number of iterations x 1,000

$$\frac{du_1}{dt} = (x_1 - x_5u_2 - x_6u_3 - x_7u_4)u_1$$
$$\frac{du_2}{dt} = (x_2 + x_5u_1 - x_8u_3)u_2 + \frac{1}{4}x_7u_4u_1$$
$$\frac{du_3}{dt} = (x_3 + x_6u_1 - x_9u_2)u_3 + \frac{1}{4}x_7u_4u_1$$
$$\frac{du_4}{dt} = (x_4 + \frac{1}{2}x_7u_1)u_4 + (x_8 + x_9)u_3u_2$$
$$u_1(t = 0) = x_{10}$$
$$u_2(t = 0) = x_{11}$$
$$u_3(t = 0) = x_{12}$$

 $u_4(t=0) = x_{13}$



Hongyu Miao, Carrie Dykes, Lisa M. Demeter, James Cavenaugh, Sung Yong Park, Alan S. Perelson, and Hulin Wu. Modeling and Estimation of Kinetic Parameters and Replicative Fitness of HIV-1 from Flow-Cytometry-Based Growth Competition Experiments. Bulletin of Mathematical Biology, 70:1749–1771, 2008.

$$\begin{aligned} \frac{du_1}{dt} &= (x_1 - x_5u_2 - x_6u_3 - x_7u_4)u_1\\ \frac{du_2}{dt} &= (x_2 + x_5u_1 - x_8u_3)u_2 + \frac{1}{4}x_7u_4u_1\\ \frac{du_3}{dt} &= (x_3 + x_6u_1 - x_9u_2)u_3 + \frac{1}{4}x_7u_4u_1\\ \frac{du_4}{dt} &= (x_4 + \frac{1}{2}x_7u_1)u_4 + (x_8 + x_9)u_3u_2\\ u_1(t = 0) &= x_{10}\\ u_2(t = 0) &= x_{11}\\ u_3(t = 0) &= x_{12}\\ u_4(t = 0) &= x_{13} \end{aligned}$$



Hongyu Miao, Carrie Dykes, Lisa M. Demeter, James Cavenaugh, Sung Yong Park, Alan S. Perelson, and Hulin Wu. Modeling and Estimation of Kinetic Parameters and Replicative Fitness of HIV-1 from Flow-Cytometry-Based Growth Competition Experiments. Bulletin of Mathematical Biology, 70:1749–1771, 2008.

Conventional Algorithm



Average Relative Parameter Estimation Error



ARE of the *i* th parameter

$$\frac{1}{N} \sum_{j=1}^{N} \frac{x_{ij} - x_i^*}{x_i^*} \times 100\%$$

Hongyu Miao, Xiaohua Xia, Alan S. Perelson, and Hulin Wu.

On Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics. SIAM REVIEW, 53(1):3–39, 2011.

Conventional Algorithm



Conventional Algorithm		New Algorithm
	Number of simulations	
	CPU time	

Conventional Algorithm		New Algorithm
131,508	Number of simulations	
	CPU time	

Conventional Algorithm		New Algorithm
131,508	Number of simulations	11,280
	CPU time	

Conventional Algorithm		New Algorithm
131,508	Number of simulations	11,280
2 hours	CPU time	

Conventional Algorithm		New Algorithm
131,508	Number of simulations	11,280
2 hours	CPU time	10 min

$$\begin{aligned} \frac{du_1}{dt} &= x_1/(1 + (p/x_2)^{x_3} + (x_4/s)^{x_5}) - x_6u_1 \\ \frac{du_2}{dt} &= x_7/(1 + (p/x_8)^{x_9} + (x_{10}/u_7)^{x_{11}}) - x_{12}u_2 \\ \frac{du_3}{dt} &= x_{13}/(1 + (p/x_{14})^{x_{15}} + (x_{16}/u_8)^{x_{17}}) - x_{18}u_3 \\ \frac{du_4}{dt} &= x_{19}/(x_{20} + u_1)u_1 - x_{21}u_4 \\ \frac{du_5}{dt} &= x_{22}/(x_{23} + u_2)u_2 - x_{24}u_5 \\ \frac{du_6}{dt} &= x_{25}/(x_{26} + u_3)u_3 - x_{27}u_6 \\ \frac{du_7}{dt} &= x_{28}/x_{29}u_4(s - u_7)/(1 + s/x_{29} - u_7/x_{30}) \\ &- x_{31}/x_{32}u_5(u_7 - u_8)/(1 + u_7/x_{32} + u_8/x_{33}) \\ \frac{du_8}{dt} &= x_{31}/x_{32}u_5(u_7 - u_8)/(1 + u_7/x_{32} + u_8/x_{33}) \\ &- x_{34}/x_{35}u_6(u_8 - p)/(1 + u_8/x_{35} + p/x_{36}) \end{aligned}$$

Carmen G. Moles, Pedro Mendes, and Julio R. Banga. Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods. Genome Research, 13:2467–2474, 2003.

$$\begin{aligned} \frac{du_1}{dt} &= x_1/(1 + (p/x_2)^{x_3} + (x_4/s)^{x_5}) - x_6u_1 \\ \frac{du_2}{dt} &= x_7/(1 + (p/x_8)^{x_9} + (x_{10}/u_7)^{x_{11}}) - x_{12}u_2 \\ \frac{du_3}{dt} &= x_{13}/(1 + (p/x_{14})^{x_{15}} + (x_{16}/u_8)^{x_{17}}) - x_{18}u_3 \\ \frac{du_4}{dt} &= x_{19}/(x_{20} + u_1)u_1 - x_{21}u_4 \\ \frac{du_5}{dt} &= x_{22}/(x_{23} + u_2)u_2 - x_{24}u_5 \\ \frac{du_6}{dt} &= x_{25}/(x_{26} + u_3)u_3 - x_{27}u_6 \\ \frac{du_7}{dt} &= x_{28}/x_{29}u_4(s - u_7)/(1 + s/x_{29} - u_7/x_{30}) \\ &- x_{31}/x_{32}u_5(u_7 - u_8)/(1 + u_7/x_{32} + u_8/x_{33}) \\ \frac{du_8}{dt} &= x_{31}/x_{32}u_5(u_7 - u_8)/(1 + u_7/x_{32} + u_8/x_{33}) \\ &- x_{34}/x_{35}u_6(u_8 - p)/(1 + u_8/x_{35} + p/x_{36}) \end{aligned}$$

Carmen G. Moles, Pedro Mendes, and Julio R. Banga. Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods. Genome Research, 13:2467–2474, 2003.



Conventional Algorithm		New Algorithm
	Number of simulations	
	CPU time	

Conventional Algorithm		New Algorithm
174,586	Number of simulations	
	CPU time	

Conventional Algorithm		New Algorithm
174,586	Number of simulations	3,167
	CPU time	

Conventional Algorithm		New Algorithm
174,586	Number of simulations	3,167
3 days	CPU time	

Conventional Algorithm		New Algorithm
174,586	Number of simulations	3,167
3 days	CPU time	1.5 hours

Conclusion

	Nonlinear Model	Computational Cost
Monte Carlo / Bootstrap method		

	Nonlinear Model	Computational Cost
Monte Carlo / Bootstrap method	Robust	


"Practical" Parameter Identifiability Analysis

	Nonlinear Model	Computational Cost
Monte Carlo / Bootstrap method	Robust	Impractically Slow.
Monte Carlo / Bootstrap method + New Algorithm		

"Practical" Parameter Identifiability Analysis



"Practical" Parameter Identifiability Analysis



Practical Parameter Identifiability Analysis



Research funded and supported by

Natural Science and Engineering Research Council of Canada

Department of Applied Mathematics, University of Waterloo