Profile likelihood to analyze the identifiability of model parameters

By Natal van Riel, Eindhoven University of Technology, the Netherlands
March 2, 2011

Systems Biology employs mathematical modeling of biological reaction networks by systems of nonlinear differential equations. Model parameters need to be estimated using experimental data. To ensure that the model predictions are reliable, it is important to infer how well model parameters can be determined. The complexity of the models in combination with the limited amount of quantitative data could result in non-identifiability of parameters. Therefore, model identifiability and uncertainty analysis are important topics in Systems Biology modeling. Here an identifiability analysis method based on the principles underlying Maximum Likelihood Estimation will be discussed. It is assumed that the reader is familiar with parameter estimation according to Maximum Likelihood Estimation and the concepts of statistical hypothesis testing and confidence intervals.

Parameter estimation: nonlinear weighted least squares

It is assumed that model parameters are estimated by fitting a model to experimental data with a least square estimation algorithm. Here time-series data is considered for which the sampling time can be non-equidistant. If data on several system variables is available, the model is fitted to all data simultaneously. The cost function is defined as

$$\chi^2(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{N}\left(\frac{y_i(t_j) - y_i(t_j \mid \theta)}{\sigma_{ij}}\right)^2 \qquad (1)$$

where $y_i(t_j)$ denotes the data-point for *i*-th observable, measured at time point t_j , and $y_i(t_j \mid \theta)$ the corresponding *i*-th model output at time t_j given the parameters θ. The errors in the experimental data set, represented as the standard deviation σ, are used as weights for each data point, quantifying their importance.

Multiple solutions that fit the data: Monte Carlo Multiple Minimization

Minimization of (1) provides a set of estimated parameters $\hat{\theta}$. If the model is not-identifiable, then this estimate is not unique. To detect possible multiple local minima a Monte Carlo Multiple Minimization (MCMM) approach can be used

$$\hat{\theta}_k = \arg\min_{\theta \geq 0} \chi^2(\theta) \quad for \quad \theta_k^0$$

in which the initial guesses of the parameters θ_k^0 are randomly chosen. To prevent that the randomly drawn initial guesses are accidentally close to each other, multiple fits with latin hypercube sampling of the initial guesses can be performed. This means, that for N_{lhs} samples, the domain of each parameter component is subdivided into N_{lhs} equally-sized intervals

(possibly using a logarithmic scale). The vector samples are drawn so that each interval for each component is chosen exactly once. Within the intervals, the parameter components are drawn uniformly distributed.

Parameter accuracy: Profile Likelihood

Profile likelihood is used to analyze the identifiability. The idea of the approach is to explore the parameter space for each parameter in the direction of the least increase in the cost function χ^2 (1). One starts with the best fit parameters and then changes one parameter (denoted with *i*) incrementally while optimizing for all the others. While performing this traversal, the other parameters are continually re-optimized, hereby tracing a path through parameter space. A profile likelihood is calculated for each parameter θ_i as the weighted residual sum of squares (WRSSs) along the path:

$$\chi^2_{PL}(\theta_i) = \arg\min_{\theta_{j\neq i}} \chi^2(\theta)$$

by re-optimization of $\chi^2(\theta)$ with respect to all parameters $\theta_{j\neq i}$, for a range of values for parameter θ_i, therefore keeping χ^2 as small as possible alongside θ_i.

In some cases, model parameters can be functionally related. These relations are referred to as *structural non-identifiabilities* and manifest themselves through a constant $\chi^2_{PL,i}$ for the involved parameters (Fig. 1). In other cases parameter values cannot be reliably inferred due to the measurement noise or a limited amount of information in the data. These parameters shall be referred to as *practically non-identifiable*.

After performing the analysis, it is important to verify that the PL covers all the acceptable solutions obtained from the MCMM. If this is not the case then this means that another local optimum apparently exists and the analysis should be repeated starting from this optimum. The profiles can then be merged afterwards.

The fact that each parameter is treated independently (performing only a 1D traversal for each parameter) makes this method efficient to compute.

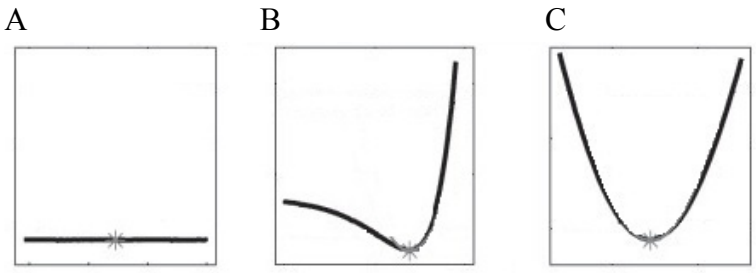

Figure 1 Three examples of typical profile likelihoods. A) A structural non-identifiable parameter is indicated by a flat PL. B) A practically non-identifiable parameter. C) An identifiable parameter shows a bounded PL, typically with a quadratic shape. The original, best fit parameter value is indicated with a *.

Profile Likelihood in Matlab

The Profile Likelihood method has been implemented in the Matlab function proL. This function calculates the profile likelihood in one direction (decreasing or increasing) of the parameter range, starting from an obtained best-fit parameter set (also referred to as the calibrated parameters).

[plPar,plRes]=proL(func,par,i,maxPar,threshold,lb,ub,Optimoptions,minStep,maxStep, minChange,maxChange,nr)

with input arguments:
- func : function handle referring to cost function for model fitting (lsqnonlin); this cost function should be defined (and implemented) by the user
- par : calibrated parameters
- i : the i-th parameter for PL calculation
- maxPar : max value of i-th parameter
- threshold : threshold - chi square distribution
- lb : lower bounds for parameter estimation (lsqnonlin)
- ub : upper bounds for parameter estimation (lsqnonlin)
- Optimoptions : options for optimization
- minStep : minimal step factor
- maxStep : maximal step factor
- minChange : minimal change of resnorm
- maxChange : maximal change of resnorm
- nr : no. of samples in profile likelihood

The output of the function:
- plPar : vector of values for i-th parameter
- plRes : vector of corresponding resnorm values (profile likelihood)

Using the default settings (Table 1):

[plPar,plRes]=proL(func,par,i)

Table 1: Default values for proL function

Algorithm settings	default values
maxPar	10*par(i); 10 times the calibrated parameter value
threshold	chi2inv(0.5,size(par));
lb	[]
ub	[]
Optimoptions	[]
minStep	0.01
maxStep	0.1
minChange	0.001
maxChange	0.05
nr	100

Example: An irreversible monomeric reaction

The chemical model for the reaction is:

$$A \xrightarrow{k} \varnothing$$

The corresponding differential equation is:

$$\frac{dy}{dt} = -ky$$

with $y = [A]$, and both the initial condition $y(0) = A_0$ and rate constant k unknown.

The unknown parameters $\theta = [k, A_0]$ are estimated by fitting the model to experimental data containing 25 time samples (data is provided in the Appendix). The optimization criterion is a SSE. To optimize the parameters Matlab lsqnonlin was used, with initial values $\theta^0 = [1,1]$, zero lower bounds for both parameters, and the default settings for the algorithm (no options specified). The experimental data and the resulting fitted model are shown in Figure 1. The resulting MLE is $\hat{\theta} = [0.074, 0.95]$ with a residual SSE of 0.49.

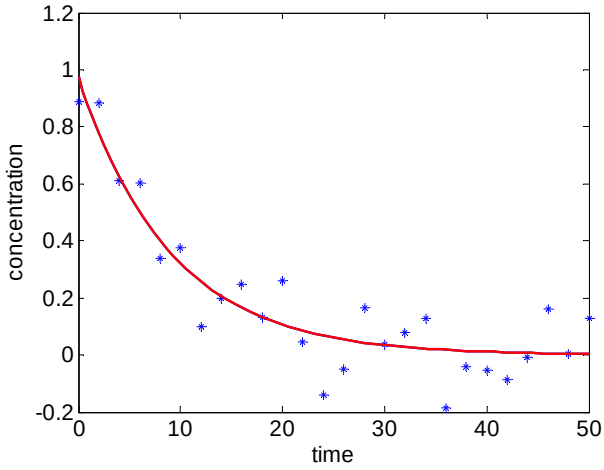

Figure 2 Experimental data and the fitted model for the irreversible monomeric reaction.

To identify the possible existence of multiple local minima a MCMM was performed starting in 20 different initial values uniformly distributed around the obtained estimate $\hat{\theta}$. A single global optimum was found (Fig. 3A, note the scale of the axes). This was confirmed with a plot of the error function (SSE, Fig. 3B).

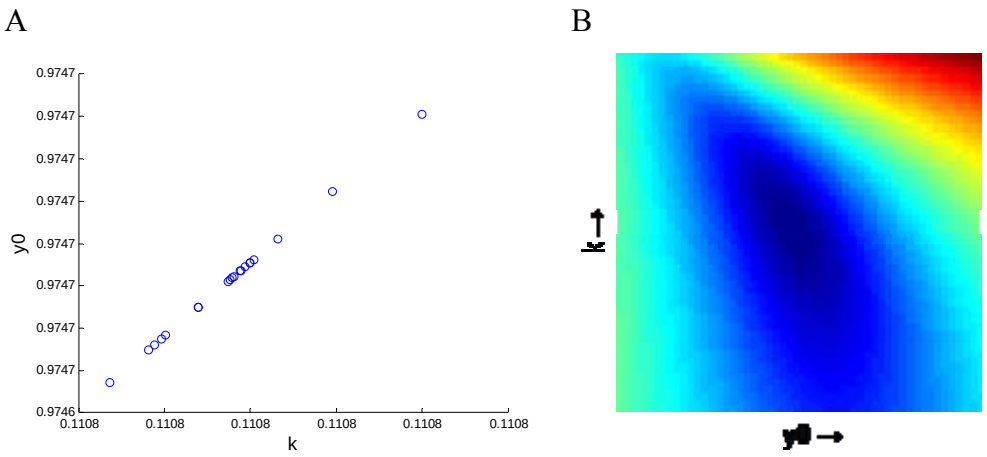

Figure 3 First step in the identifiability analysis of an irreversible monomeric reaction A) MCMM result, note the scales of the axes. B) Error function plot.

As a second step in the identifiability analysis the Profile Likelihoods were determined for the two parameters. Fig. 4 shows both parameters are identifiable. The PL for y_0 shows a symmetric, quadratic profile, whereas the PL for k is somewhat 'flattened' in the increasing direction, which is also visible in the error function plot (Fig. 3B).
The Matlab code to calculate the Profile Likelihood for the irreversible monomeric reaction using the proL function can be found in the Appendix.

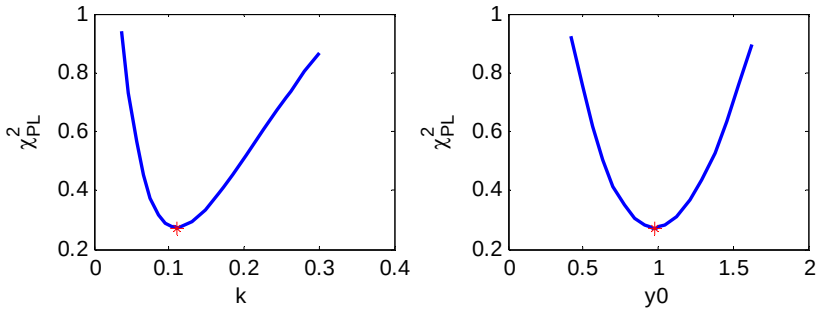

Figure 4 Second step in the identifiability analysis of an irreversible monomeric reaction: Profile Likelihood.

Confidence intervals of estimated parameters
Confidence intervals of estimated parameters can be derived using a threshold in the likelihood. An approximate likelihood-based confidence region for the parameter is the set of all values of θ such that:

$$\left\{\theta \mid \chi^2(\theta_{PL}) - \chi^2(\hat{\theta}) < threshold\right\}$$

The threshold is chosen by the modeler. Theory to select a suitable threshold is related to statistical hypothesis testing.

Likelihood Ratio In statistics, a likelihood ratio (LR) test is a statistical test used to compare the fit of two models, one of which (the null model M_0) is a special case of the other (the alternative model M_1). In general, a model with more free parameters will fit at least as well (have a greater log-likelihood) than a model with less parameter. Whether it fits significantly better and should thus be preferred is determined by analyzing the difference in the likelihoods. The test is based on the likelihood ratio, which expresses how many times more likely the data are under one model than the other. This likelihood ratio, or equivalently its logarithm, can then be compared to a critical value (or used to compute a p-value) to decide whether to reject the null model in favor of the alternative model.
Each of the two competing models, the null model and the alternative model, is separately fitted to the data and the log-likelihood recorded. The test statistic (often denoted by D) is twice the difference in these log-likelihoods:

$$D = -2\ln\left(\frac{L(M_0)}{L(M_1)}\right)$$

Nested models and χ^2 distribution Where the null hypothesis represents a special case of the alternative hypothesis, the probability distribution of the test statistic is approximately a chi-squared distribution χ^2_p with degrees of freedom p equal to the difference in the number of free parameters of models 0 and 1, the null model and the alternative model, respectively. This requires nested models, that is: models in which the more complex one can be transformed into the simpler model by imposing linear constraints on the parameters.

In case of Profile Likelihood, the null hypothesis is the model in which one parameter is fixed to a value that is different from its best fit value $M(\theta_{PL})$. Model $M(\theta_{PL})$ results from the best fitting model $M(\hat{\theta})$ by imposing a constraint on one of the parameters and therefore the two model are nested and their LR is approximately distributed according to a χ^2_p distribution with $p = 1$. In PL analysis one parameter is fixed at a new value and therefore the χ^2 distribution of the LR has one degree of freedom.

Significance level Finally, a desired significance level α has to be chosen. Confidence Intervals (CIs) [1] are computed for the region of the PL for which holds:

$$D = -2\ln\left(\frac{L(\theta_{PL})}{L(\hat{\theta})}\right) \leq \chi^2_{1-\alpha,1} \qquad (2)$$

Confidence bounds for each parameter are then obtained as the parameter values where (9) becomes an equality for the first time.

[1] In statistics textbooks, confidence intervals are typically introduced for a standard Normal distribution N(0,1) and a 95% confidence level is considered. N(0,1) places 95% of probability between −1.96 and 1.96.

In case $\chi^2_{p=1}$ is used to define the threshold for the CIs (degrees of freedom *df* =1), then (2) gives *pointwise confidence intervals* that hold individually for each parameter.
Another choice of *df* is *df* = #*θ* being the number of parameters which gives *simultaneous confidence intervals* that hold jointly for all parameters.
The chi-square quantile function (inverse of the cumulative distribution function) in Matlab:
chi2inv(0.5,size(par));
This is the value that is by default used in proL as threshold criterion to stop the algorithm (Table 1).

In case of structural non-identifiable parameters no parameter confidence bound can be computed for such a parameter since a perturbation of one parameter can be negated by varying another.

References

1. Meeker,W. and Escobar,L. (1995) Teaching about approximate confidence regions based on maximum likelihood estimation. Am. Stat., 49, 48–53.
2. Murphy S.A. and A.W.van der Vaart (2000) On Profile Likelihood. J Am Stat Assoc 95(450): 449-465
3. Raue A., C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmuller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics, 25(15): 1923-1929, 2009.
4. Vanlier J., C.A. Tiemann, P.A.J. Hilbers and N.A.W. van Riel (2012) An Integrated Strategy for Prediction Uncertainty. Bioinformatics. 2012 Feb 21. [Epub ahead of print]

Appendix A Matlab files to calculate the PL for the irreversible reaction using proL function

Data

```
data = [
         0    1.0538
    2.0000    1.0021
    4.0000    0.4457
    6.0000    0.6357
    8.0000    0.4818
   10.0000    0.2378
   12.0000    0.2579
   14.0000    0.2813
   16.0000    0.5600
   18.0000    0.4425
   20.0000    0.0006
   22.0000    0.4143
   24.0000    0.1634
   26.0000    0.0681
   28.0000    0.1324
   30.0000    0.0294
   32.0000    0.0284
   34.0000    0.1824
   36.0000    0.1683
   38.0000    0.1641
   40.0000    0.0855
   42.0000   -0.1058
   44.0000    0.0840
   46.0000    0.1731
   48.0000    0.0571
   50.0000    0.1102];

texp = data(:,1); ydata = data(:,2);
```

Fit function: decay_fit.m

```
function e = decay_fit(p, texp, ydata)

k = p(1);
y0 = p(2);

[t y] = ode45(@decay_ode, texp, y0, [], k);

e = ydata(:)-y(:);
```

ODE model function: decay_ode.m

```
function dydt = decay_ode(t, y, k)
```

```
dydt(1) = -k*y(1);

dydt=dydt(:);   %column
```

Script to calculate PL

```
parname = {'k','y0'};
p = [0.074, 0.95];
resnorm = 0.491;

%% Profile Likelihood
func = @(p)decay_fit(p,texp,ydata); %cost function for model fitting
thresh = chi2inv(0.5,length(p)); %threshold - chi square distribution
lb = [0 0]; %lower bounds for parameter estimation (lsqnonlin)
ub = []; %upper bounds for parameter estimation (lsqnonlin)
optimoptions = []; %options for optimization (lsqnonlin)
minStep = 0.01;     %minimal step factor
maxStep = 0.1;      %maximal step factor
minChange = 0.01;   %minimal change of resnorm
maxChange = 1;      %maximal change of resnorm
nr = 10;            %no. of samples in profile likelihood

% PL for parameters in the increasing direction
maxPar = [2 2];     %max values parameters
for i=1:length(p)
[plPar,plRes]=proL(func,p,i,maxPar(i),thresh,lb,ub,optimoptions,minStep,max
Step,minChange,maxChange,nr);

    parInc{i} = plPar;
    resInc{i} = plRes;
end

% PL for parameters in the decreasing direction
maxPar = 0.1*p;
for i=1:length(p)
[plPar,plRes]=proL(func,p,i,maxPar(i),thresh,lb,ub,optimoptions,minStep,max
Step,minChange,maxChange,nr);

    parDec{i} = fliplr(plPar);
    resDec{i} = fliplr(plRes);
end

figure
for i=1:length(p)
    subplot(2,2,i)
    plot([parDec{i} parInc{i}], [resDec{i} resInc{i}], 'Linewidth',2); hold
on
    plot(p(i),resnorm,'*r')
    plot([parDec{i}(1) parInc{i}(end)], [thresh thresh], '--k')
```

```
    xlabel(parname(i));ylabel('\chi^2_{PL}');
end
```