THE LEAST SQARES METHOD APPLIED TO THE LIFE SCIENCE MODELS

ESTIMAREA PARAMETRILOR IMPLICAȚI ÎN MODELE DE CREŞTERE

ALDEA FLORICA, POP IOANA, STANCA LIANA, MICULA MARIA
University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca

Abstract: *When we model a phenomenon is important to find an analytical expression. Usually this expression involve several parameters that can be determined using least squares method or category of its. In this paper we focus on application of this method in cases when direct implementation do not work. For that we use liberalization or Levenberg-Marquardt algorithm.*

Rezumat: *In ştiinţele vieţii explicarea diverselor fenomene se bazează pe elaborarea unor modele matematice. Acestea sunt rezultatul transpunerii analitice a legilor fizice, chimice sau biologice care guverneaza fenomenele respective. Practic, există posibilitatea achiziţiei unor date experimentale pe baza cărora se vor determina parametrii care definesc modelul matematic. Metoda celor mai mici pătrate se bazează pe distribuţia Gauss de verosimilitate a datelor experimentale. Această funcţie îşi atinge maximul atunci când suma pătratelor diferenţelor dintre valorile experimentale şi valorile funcţiei date de modelul matematic este minimă. Se prezintă două metode de determinare a parametrilor din model.*

INTRODUCTION

In applied science the mathematical model of studied phenomenon plays an important role. Mathematical model is important because offers the possibility to save time and money and gives appropriate information about the studied phenomenon. Such model, in general, contains a number of parameters. A parameter is identified by the fact that it takes the same fixed value. Often, there is some initial information about the parameters values based on scientific experience and/or the values reported in the scientific literature. Even for the measurable parameters one is often confronted with a range of possible values due to the measurements errors or due to real difference for different circumstances. One of the frequent methods to determine the parameters involve in different mathematical model is the least squares method and its category.

MATERIAL AND METHOD

If we use a set of N experimental points $a_i, i = 1,..., N$ in order to determine N quantities $y_i, i = 1,..., N$ than each observational error is given by

$\varepsilon_i = a_i - y_i, i = 1,..., N$ and satisfy Gauss distribution [1]. The differential of observational errors are $d\varepsilon_i = da_i, i = 1,..., N$ and the function

$$\varphi(a_i - y_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{a_i - y_i}{\sigma}\right)^2} da_i \quad (i = 1,..., N) \tag{1}$$

represents the probability that experimental values are between a_i and $a_i + da_i$. In other words function (1) gives the probabilities with that we obtain value a_i when we measure y_i. The precision of each measurement is given by the dispersion $\sigma_i, i = 1,..., N$. The probability *P* is given by

$$P = \frac{1}{\left(\sqrt{2\pi}\right)^N} \frac{1}{\sigma_1} \frac{1}{\sigma_2} \cdots \frac{1}{\sigma_N} e^{-\frac{1}{2}\left[\left(\frac{a_1 - y_1}{\sigma_1}\right)^2 + \left(\frac{a_2 - y_2}{\sigma_2}\right)^2 + \cdots + \left(\frac{a_N - y_N}{\sigma_N}\right)^2\right]} da_1 da_2 \cdots da_N \tag{2}$$

it represent the probability as simultaneous for each quantities y_i to obtain a_i value. Analysing relation (2) we have that probability *P* is maximized when

$$S = \sum_{i=1}^{N} \frac{(a_i - y_i)^2}{\sigma^2} \quad \rightarrow \quad \min \tag{3}$$

The minimum of (3) is obtained in point that the partial derivatives of S respect to $a_i, i = 1,..., N$ vanish. We consider that mathematical model depends of parameters $\alpha, \beta, \gamma, ...$ that must be determined using experimental points $a_i, i = 1,..., N$. In literature there are distinguish tow different cases due to minimization of (3). The common case when we have a linear and nonhomogenous system with unknown $\alpha, \beta, \gamma, ...$ and the most delicate case when the system is nonlinear. For the second situation there are three different methods to determine the parameters of interest [2]. Each methods use the assumption that we have a trial solution.

The first method used the Taylor series expansion of function *S* in the neighbourhood of trail solution in order to linearize *S*. At the end we obtained a solution $\alpha+\Delta\alpha, \beta+\Delta\beta, \ldots \gamma+\Delta\gamma$. Using an iterative procedure is possible to determine an acceptable solution if *S* reaches its minimum.

The second method to obtain the values $\alpha, \beta, \gamma, ...$ such that S reaches its minimum was developed by Nelder and Mead [3] named simplex (but has nothing in common with simplex algorithm from linear programming).

The last method is based on previous tow and was developed by Levenberg and Marquardt [5]

RESULTS AND DISCUSSIONS

In this section we discuses two examples that will illustrate the least squares method in the case when the model will be linearlized such that the least squares method can be applied (Application 1) and the other example uses a functional

that cannot be linearlized and will be used the Levenberg-Marquard algorithm (Application 2).

Application 1 The decreasing in time of a number of biological cells is given by

$$N(t) = N_0 e^{-\lambda t} \tag{4}$$

where N_0 is the initial number of cells and λ gives the death probability for cells. Using the least squares method we find out N_0 and λ parameters. Linearizing the relation (4) we obtain

$$\ln N(t) = -\lambda t + \ln N_0 \tag{5}$$

and denoting $Y(t) = \ln N(t)$, $\ln N_0 = \gamma$ we have

$$Y(t) = -\lambda t + \gamma \tag{6}$$

In Fig. 1 are given the experimental values and theirs approximation with the best values for parameters, N_0=35.45 cells and $\lambda = 0.029$ 1/hour.

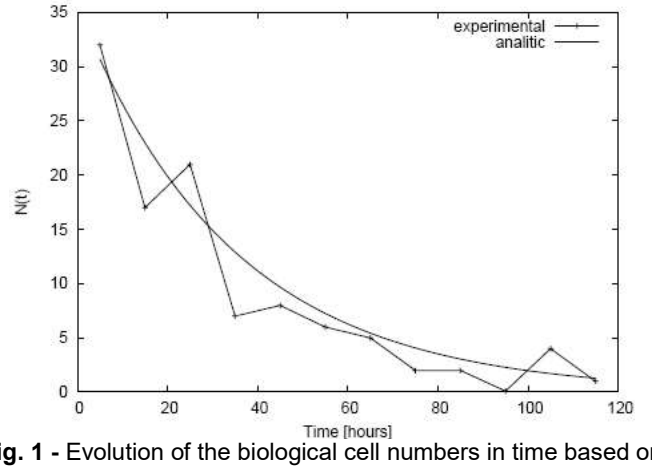

Fig. 1 - Evolution of the biological cell numbers in time based on N(t)=35.45exp(-35.45 t) relation

The results from Fig. 1 are based on following considerations [3]. For the initial set of data (t_i, y_i), $i=1, \ldots, N$ was accepted that the errors are described by the normal distribution. After liniarization the set of experimental points are not described by Gauss distribution such that least square method cannot find the best parameters for the model. This problem can be avoided using global weights.
When we apply the least squares method we have:

$$\sum_{i=1}^{N} \left(\underbrace{N_0 e^{-\lambda t_i} - y_i^{\exp}}_{\Delta N_i} \right)^2 \rightarrow \sum_{i=1}^{N} w_i \left(\underbrace{\ln N(t_i) - \ln y_i^{\exp}}_{\Delta Y_i} \right)^2 \tag{7}$$

By taking into account $\frac{\Delta Y_i}{\Delta N_i} \cong \left.\frac{dY}{dN}\right|_{t=t_i} = \frac{1}{y_i^{\exp}}$ that means weights is given by $w_i = \left(y_i^{\exp}\right)^2$.

The importance of these considerations is illustrated in the next example. Let consider another set of experimental data $(x_i, y_i), \quad i = 1,2,\cdots,N$ that can be approximated by Gauss distribution. In this case the weights are the same as in previous application. Fig. 2 represents the least squares method with weights and without.

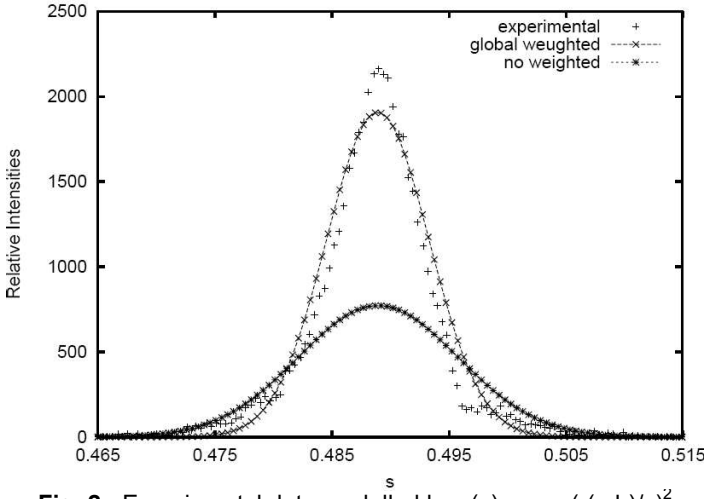

Fig. 2 - Experimental data modelled by $y(s)=a\exp(-(s-b)/c)^2$

Application 2. We consider the experimental set of points $(x_i, y_i, z_i), i = 1,...,N$, The mathematical model that we consider is given by

$$h(x,y) = \sqrt{r^2 - (x-x_0)^{2.2} - (y-y_0)^{1.8}} + z_0 \qquad (5)$$

The interest parameters are x_0, y_0, z_0 and r. We want to determine the best values for the parameters but we cannot use direct least squares method because we cannot linearized. In this case we used the Levenberg Marquardt algorithm that is implemented in ***Gnuplot*** software.

For this application we have implemented the following code sequence:

```
print "\n The demonstration of 3d fitting by using Lavemberg Marquard algorithm."
print "hemisphr.dat contains experimental data on a hemisphere "print "It takes many iterations, so we limit FIT_MAXITER to 50."
#: Analytical model is given by relation
h(x,y) = sqrt(r*r - (abs(x-x0))**2.2 - (abs(y-y0))**1.8) + z0
```

```
# Trial (initial) solution of the model
x0 = 0.1
y0 = 0.2
z0 = 0.3
r=0.5
FIT_MAXITER=50
set xlabel "x"
set ylabel "y"
set zlabel "z"
set xtics -1,0.5
set ytics -1,0.5
set ztics -1,0.5
```

#Initial solution of the model' splot 'hemisphr.dat' using 1:2:3, h(x,y) pause -1 "(-> return)" fit h(x,y) 'hemisphr.dat' using 1:2:3:(1) via r, x0, y0, z0 #Final solution of the model and its 3D reprezentation' splot 'hemisphr.dat' using 1:2:3, h(x,y) pause -1 "(This is an exemple of Levemberg Marquardt algorithm). (-> return)"

After we have executed the set of commands we obtain the parameter: x_0=-0.00031, y_0=0.000599, z_0=0.002538, and r=1.00193. In Fig. 3 are given the initial set of points and the result after global approximation is given in Fig. 4

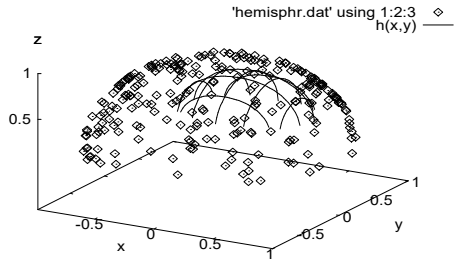

Fig. 3 - Experimental data and trial solution of the mathematical model

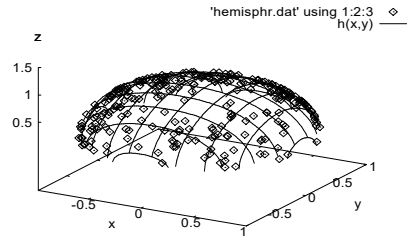

Fig. 4 - Experimental data and the best solution of the mathematical model based on Levenberg and Marquardt algorithm

CONCLUSIONS

In case that mathematical model is linear there are no problems to apply direct the least square method. In the other cases our conclusions are:

- There are several software products that make easier this problem and we mention Maple software. In the cases when the model can be linearized is important to use weights because other way after liniarization the experimental set of point do not respect normal distribution otherwise are obtained irrelevant results.
- If the model cannot be linearized is mandatory to use some methods to find parameters. Is possible to use in Taylor series expansion if the trail solution is not to far from the best solution.
- The Levenberg Marquardt algorithm can be used with good results even the trail solution is far from solution. Many scientists use this algorithm due to its remarkable results

REFERENCES

1. **Aldea Florica, 2006** - *Matematici aplicate in stiintele agricole si silvice*. Ed. Risoprint .
2. **Aldea N., Aldea Florica, 2001** - *Tehnici de analiza a semnalelor fizico-chimice*. Ed. Risoprint.
3. **Estera J. P., Duncan W. M., Slaughter S.R., 1992 -** *Lineshape Analysis for Optical Transition in III-V Semiconductors Using a Sequential Simplex Procedure***.** Comp. In Physiscs, 6 (1992) 360.
4. **De Levie R.** - *When, Why and How to Use Weighted Least Squares*. J. Chem. Educ., vol.63
5. **Marquardt D.**, **1963 -** *An algorithm for least-suares estimation of nonlinear parameters*. J. Soc. Indust. Appl. Math., vol 11, no2.
6. **Williams T., Kelly C.** - *Gnuplot An Interactive plotting program.* prepared by D. Crawford version 3.7