The Population Predicting Based on the Curve Fitting Least Square Method

Lili HE[1,a],Zhao JIN[2*,b]

[1]College of Information Science and Electronic Technology JIA MU SI University,China

[2]JIA MU SI Aviation Forest Fire,China

[a]jmshelili@163.com,[b]12316036@qq.com

Keywords: Curve Fitting; Malthusian model; prediction model

Abstract. In various scientific experiments, experimenter always get some discrete data, to estimate the function with the original rules from these data, interpolation is a method for dealing with this kind of problem.But when the discrete data is too much,it can be caused serious distortion which is called Runge phenomenon at the interval endpoints.To avoid，the Curve Fitting Least Square method is a kind of optimal solution. The Curve Fitting Least Square method is used to be dealing with structure prediction model etc.In this paper, Malthusian model is the method of population predicting.In Malthusian model,the unknown variable is ascertained by the Curve Fitting Least Square method.

Introduction

In the scientific experiments or the statistical research,always get some discrete data from experiment or sampling,recorded as $(x_i, y_i)(i = 0,\cdots n)$.Try to structure the function $p(x)$ by these discrete data to replace the original function $f(x)$.The feature of commonly used interpolation method is the fitting curve crossing those interpolation points,such as Largange,Newton,Hermite and Cubic Spline etc.Although some data may be somewhat distorted by the experiment ,or defective data, interpolation method can take those errors to $p(x)$,and given the inevitable adverse effects. The Curve Fitting Least Square method is not pass by every ineterpolation point,which is pass through those point by fitting thought,to get the change trend of curve and isolate causal,and reflect the objective fact.

The principle of the Curve Fitting Least Square method

If the experiment dataset is $(x_i, y_i)(i = 0,\cdots n)$,and the fitting function is $p(x)$,then the deviation which is maked from every interpolation point is $\delta_i = \left|p(x_i) - y_i\right|(i = 0,\cdots n)$. $p(x)$ can express the original function $f(x)$,and make δ_i minimum. Usually use these standards as a measure of the deviation:

$$\|\delta_i\|_1 = \sum_{i=0}^{n}|\delta_i|,\quad \|\delta_i\|_2 = \left(\sum_{i=0}^{n}\delta_i^2\right)^{\frac{1}{2}},\quad \|\delta_i\|_\infty = \max_{0\le i\le n}|\delta_i|$$

The Curve Fitting Least Square method use minimum 2-norm.

More specifically, for a group of experiment data $(x_i, y_i)(i = 0,\cdots n)$,try to struct a function as $p(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots a_m\varphi_m(x) = \sum_{j=0}^{m} a_j\varphi_j(x), \varphi(x)$ is the basis function in it, When $\|\delta_i\|_2^2 = \sum_{i=0}^{n}\delta_i^2 = \sum_{i=0}^{n}\left(p(x_i) - y_i\right)^2$ is least, $(a_0, a_1, \cdots a_m)^T$ is the solution.

Multi-variable function is $F(a_0, a_1, \cdots a_n) = \|\delta_i\|_2^{2^2} = \sum_{i=0}^{n}\left(\sum_{j=0}^{m} a_j \varphi_j(x_i) - y_i\right)$,let $\frac{\partial F}{\partial a_k} = 0$

$(k = 0,1\cdots, m)$,then $\frac{\partial F}{\partial a_k} = \sum_{i=0}^{n}\left(2\left(\sum_{j=0}^{m} a_j \varphi_j(x_i) - y_i\right)\varphi_k(x_i)\right) = 2\sum_{j=0}^{m}\sum_{i=0}^{n} a_j \varphi_j(x_i)\varphi_k(x_i) - 2\sum_{i=0}^{n} y_i \varphi_k(x_i) = 0$

This is the final result which the above fomular after deducing.

$$\sum_{j=0}^{m}\sum_{i=0}^{n} a_j \varphi_j(x_i)\varphi_k(x_i) = \sum_{i=0}^{n} y_i \varphi_k(x_i)\ (k = 0,1\cdots, m).$$

Let $(\varphi_j, \varphi_k) = \sum_{i=0}^{n} \varphi_j(x_i)\varphi_k(x_i)$, $(f, \varphi_k) = \sum_{i=0}^{n} y_i \varphi_k(x_i) = d_k$,then this fomula is $\sum_{j=0}^{m} (\varphi_k, \varphi_j) a_j = d_k$ $(k = 0,1\cdots, m)$.And turn it to the matrix:

$$\begin{pmatrix} (\varphi_0,\varphi_0) & (\varphi_0,\varphi_1) & \cdots & (\varphi_0,\varphi_m) \\ (\varphi_1,\varphi_0) & (\varphi_1,\varphi_1) & \cdots & (\varphi_1,\varphi_m) \\ \vdots & \vdots & \vdots & \vdots \\ (\varphi_m,\varphi_0) & (\varphi_m,\varphi_1) & \cdots & (\varphi_m,\varphi_m) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_m \end{pmatrix}.$$

This is a linear system of equations which coefficient is $a_j(k = 0,1\cdots, m)$,and it also called canonical equation.For the vector $(\varphi_0, \varphi_1, \cdots \varphi_m)$ is linearly independent,then this linear system of equations' s determinant is not zero,so this linear system of equations have have only one solution.

Algorithm design

Curve Fitting Least Square method consists of two parts,first,get the augmented matrix of the linear system of equations's determinant.Second,get the solution of the linear system of equations.Use the Gauss method to get the solution.The specific algorithm is in the figures 1.

- $i = 0$ to m
  - $j = 0$ to m
    - $a_{ij} = \sum_{k=0}^{n} x_k^{i+j}$
  - $a_{im+1} = \sum_{k=0}^{n} y_k x_k^{i}$
- $k = 0$ to $m-1$
  - $i = k+1$ to m
    - $t = -a_{ik} / a_{kk}$
    - $j = k+1$ to $m+1$
      - $a_{ij} = a_{ij} + a_{kj} \times t$
- $i = m$ to 0
  - $a_{i,m+1} = \left(a_{i,m+1} - \sum_{j=i+1}^{m} a_{ij} \times a_{j,m+1}\right) / a_{ii}$

Fig.1: the algorithm description

In figure 1-1, n is the high limit of those interpolation points' s subscript, m is the highest power of empiric formulas,and X 、Y are the Domain and range of interpolation points.

Malthusian model

Malthusian model is a model about population growth.Let total population is $N(x)$,for the theory of population as Malthusian model, use the exponential function as $N(x)=e^{a_0+a_1x}$,and in this expression x is the year.And make the log of this expression,we can get the equation as $y(x)=\ln N(x)=a_0+a_1x$ .In the equation, a_0 and a_1 are unknown,they need fitting from those ineterpolation points,for the empiric formulas $y(x)$ is linear,and the quantity of ineterpolation points is more than 2(≥ 2),so use the Curve Fitting Least Square method to struct fitting function is appropriate.

For example,Chease population is known int the past ten years as table 1.

Table 1: Chease population in last ten years

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Quantity(hundred million)	12.9988	13.0756	13.1448	13.2129	13.2802	13.3474	13.4100	13.4735	13.5404	13.6072

From the Curve Fitting Least Square method,get the linear system of equations about a_0 and a_1:

$$\begin{pmatrix} 10 & 20085 \\ 20085 & 4.0340800 \end{pmatrix}\begin{pmatrix} a_0 \\ a_1 \end{pmatrix}=\begin{pmatrix} 25.8834 \\ 51987.3 \end{pmatrix}$$

After solving,the fitting function is $y(x)=-7.50297+0.0050243\times x$ .And put the 2014 to the Malthusian model,then the forecast is $e^{-7.50297+0.0050243\times 2014}\approx 13.6806$ hundred million.It can be expressed in figure.2 and figure 3,and in the figure.2 it is the rule in the domain,and in the figure 3it is the rule in the past ten years.It is log function rule in all the domain,but in the past ten years,it is resembling a straight line.From the figure 3,it can be found that fitting curve is not pass on every point which is pass through.

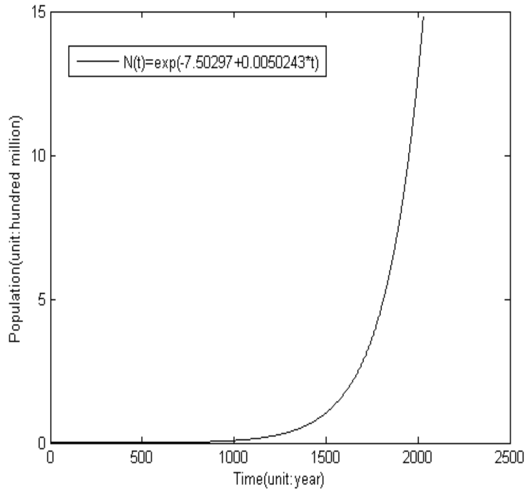

Fig.2: fitting function in the domain

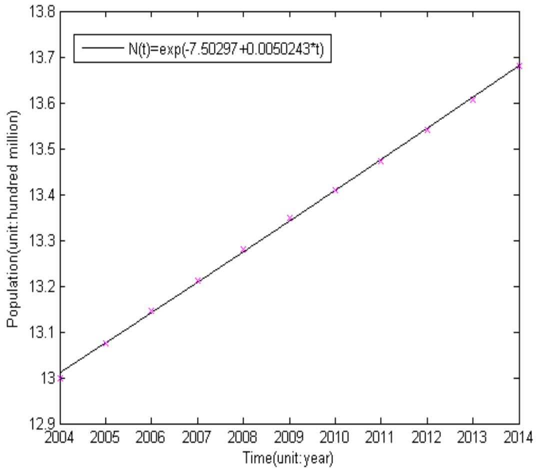

Fig.3: fitting function in the past ten years

Conclusion

The Malthusian model which is used to be forcasting population is mature and widely used.It would be assistant decision-making about decisions of a nation or area scientifically and reasonably.And the Curve Fitting Least Square method is excellent in the prediction field.Both of these method could reach the changes of the population veritably.

Acknowledgement

This work is supported by the JIA MU SI University science and technology research project(No.L2013-075)

*Corresponding author.

References

[1] YU Hongjun.Two Geometric Methods of Curve-fitting.Science Technology and Engineering.Vol.12 N0.16 Jun.2012

[2] Fu Yanru.Rearch on Actual Measurement on Matlab Curve Fitting.Journal of Jinlin Normal University.No.2.May.2010

[3] Yang Lijian,Sun Xiao,Gao Songwei.Application of curve fitting in signal reconstruction of magnetic flux leakage inner detection for pipeline.Journal of Shenyang University of Technology.Vol.34 No.6 Nov.2012

[4] Bian Huanqing,Xia letian.The Application of Gray Markov Model in Population Forcasting.Mathematics in Practice and Theory.Vol42 No.7.Apr.2012

[5] He Lili.Least Squares Algorithm Desgin and C++ Realization of Curve Fitting.Journal of Hunan Post and Telecommunication College.Vol.13 No.1.Mar.2014

[6] Ren Qiang,Hou Dadao.Stochastic Model for Population Forecast:Based on Leslie Matrix and ARMA Model.Population Research.Vol35 No.2.March 2011

[7] Hu Qingwan.Curve Fitting by Curve Fitting Toolbox of Matlab.Computer Knowledge and Technology.Vol.6 No.21.July 2010

[8] Liu Nian,Du Qiuxia,Liu Gang,Yu Peiyi.Research and Application of Curve-Fitting Based on Function Combination.Computer and Information Technology.Vol.21 No.4.Aug.2013

[9] Xu Linzhou,Zhang Xinhua,Song Yuan.Bearings-only Target Para meter Estim ation A Lgorithm for Stationary Single-Station based on Curve-fit.Fire Control and Comm and Control.Vol35.Aug 2010

[10]Zeng Jinfang,Teng Zhaosheng,Application of information entropy in curve fitting recognition.Journal of Electronic Measurement and Instrument.Vol26 No.2.Febr 2012