DhakaUniv. J. Sci. 65(2): 103-105, 2017 (July)

A Procedure to Fit an Interpolating Curve to a Set of Logistic Data

Md. Shariful Islam[1], Mir Shariful Islam[2,*], A.F.M. Khodadad Khan[3] and Md. Zavid Iqbal Bangalee[4]

[1]*Department of Mathematics, Texas Christian University, Fort worth Texas– 76129, USA*

[2]*Department of Oceanography, Dhaka University, Dhaka-1000, Bangladesh*

[3]*Independent University, Dhaka-1212, Bangladesh*

[4]*Department of Applied Mathematics, Dhaka University, Dhaka-1000, Bangladesh*

Abstract

Logistic dynamics are frequently encountered in real life problems, especially in population dynamics. Data showing an appearance to follow logistic model may be interpolated by standards methods in numerical analysis. In this paper we discuss a method to fit a curve to such data using the intrinsic analytic properties of the data in terms of least square method and graphic tools in the environment of Mathematica.

Keywords: Logistics Model, Least Square Approximation, Mathematica.

I. Introduction

Consider the differential equation[1] $\frac{dy}{dt} = ay - cy^2$ where a and c are positive constants. This ordinary differential equation is used to describe population dynamics taking overcrowding into an account as an improvement of Malthus Model. Putting $r = a$ and $k = a/c$ we obtain the logistic equation[2] $\frac{dy}{dt} = ry\left(1 - \frac{y}{k}\right)$. The constant r is called the intrinsic growth constant and k is the saturation constant. The solution to this ordinary differential equation is given by $y(t) = \frac{k}{1+Ce^{-rt}}$. Any real life time dependent data that resembles the graph of the solution is called logistic data. Notice that $y(t) \to k$ as $t \to \infty$, since $e^{-rt} \to 0$.

II. Mathematical Formulation

Let a given set of data $S = \{(t_i, y_i): i = 1,2, \dots \dots \dots, n\}$. Assume that the data, if plotted gives an impression to follow some solution of a logistic equation[3]. We use the logistic equation of the form

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

where, t is an independent variable, $y = y(t)$ is the value of y at t and r and k are positive parameters. Using separation of variables, we can solve the differential equation (1) as

$$y(t) = \frac{k}{1+Ce^{-rt}} \quad (2)$$

where C is an arbitrary constant. Differentiating and simplifying (2), we may show

$$\frac{d^2y}{dt^2} = \frac{Ckr^2e^{rt}(C-e^{rt})}{(C+e^{rt})^3} \quad (3)$$

The graph of $y(t)$ has a point of inflection[4] when the second derivative in (3) vanishes. Assuming the given data shows a pattern having an inflexion at $t = t_0$.

i.e., $\frac{d^2y}{dt^2} = 0$ at $t = t_0$.

Thus from equation (3), $C - e^{rt_0} = 0$

$\Longrightarrow C = e^{rt_0}$, where t_0 is the time at which inflection occurs.

Then the solution (2) becomes

$$y(t) = \frac{k}{1+e^{-r(t-t_0)}} \quad (4)$$

Let $h(t) = \frac{1}{1+e^{-r(t-t_0)}}$ (5)

From (4) we get, $y(t) = kh(t)$ (6)

Computing the sum of the square of the errors occurred in approximating the data set S from the equation (6).

We get, $E_r = \sum_{i=1}^{n}(y(t_i) - y_i)^2$

$$= (y(t_1) - y_1)^2 + \cdots + (y(t_n) - y_n)^2$$

$$= (kh(t_1) - y_1)^2 + \cdots + (kh(t_n) - y_n)^2$$

$$= \|< kh(t_1) - y_1, \dots, kh(t_n) - y_n >\|^2$$

$$= \|k < h(t_1), \dots, h(t_n) > - < y_1, \dots, y_n >\|^2$$

$$= \|kH - M\|^2$$

where $H =< h(t_1), \dots, h(t_n) >$ and $M =< y_1, \dots, y_n >$. The vectors H and M are defined according to the definition given by Paul R. Holmas[5].

$$E_r = \|kH - M\|^2 \quad (7)$$

$$=< kH - M, kH - M >$$

$$= k^2 < H, H > -2k < H, M > + < M, M > \quad (8)$$

Clearly, E_r contains three parameters namely k, r, t_0. Minimum error occurs when

$\frac{\partial E_r}{\partial k} = 0, \frac{\partial E_r}{\partial r} = 0,$ and $\frac{\partial E_r}{\partial t_0} = 0$.

From $\frac{\partial E_r}{\partial k} = 0$

$$\Longrightarrow 2k < H, H > -2 < H, M > = 0$$

$$\Longrightarrow k = \frac{<H,M>}{<H,H>} \quad (9)$$

Putting the value of k in (8), we get $E_r =< M, M > - \frac{<H,M>^2}{<H,H>}$ (10)

Equation (10) contains just two parameters r, t_0 after elimination of k. We will now plot the graph E_r , as defined by equation (10), as a function of r and t_0 , the graph is a

* Author for correspondence. e-mail:mirsharifuldu77@yahoo.com

surface in R^3 and we locate the minimum point on this "error surface". We begin by defining a domain in the rt_0 plane, it is a simple matter to determine an interval containing t_0, as it must lie in the range of the given t data. Thus, considering the data in S, the inflection point occurs at t_0 , where t_0 is some value such that $t_1 \leq t_0 \leq t_2$. A tricky project is to determine a likely range for the reproductive growth rate r. However, from equation (4) $y'(t) = \frac{kre^{-r(t-t_0)}}{(1+e^{-r(t-t_0)})^2}$

at $t = t_0$,we get

$$y'(t_0) = \frac{kr}{4} \Rightarrow r = \frac{4\acute{y}(t_0)}{k} \qquad (11)$$

From the list plot of the data set S, we will locate the position of the inflection point roughly and determine the gradient in the neighborhood of the assumed point of inflection by calculating the slope of the line passing through the points immediately following and preceding the point of inflection. Using gradient and taking k as the maximum in the given data, since $\lim_{t\to\infty} y(t) = k$, we determine $r = r_0$. Thus it is natural that r, falls between an interval centered at r_0, i.e.,$|r - r_0| \leq l$, l can be suitably chosen from the gradients in the given data. We recommend $\frac{4Miny'(t)}{\text{Max}\,(y(t))} < r < \frac{4Maxy'(t)}{Max(y(t))}$ by using equation (11). For convenience, it may be adjusted suitably as long as it contains r_0. Then we plot the error surface in the domain $t_1 \leq t_0 \leq t_2$ and $|r - r_0| \leq l$. From the plot, we locate the position of minimum error. Say minimum error occurs at r and $t = t_0$ .Mathematica[6] becomes handy in plotting the error surface and a suitable contour plot to circumvent the co-ordinate of the minimum error with a very small closed contour (Label curves of the error function). Using r and t_0 from (9) we calculate k.

Hence, we get the logistic equation from (4) as

$$y(t) = \frac{k}{1 + e^{-r(t-t_0)}}$$

III. Results and Discussion

Now we will follow the above procedure to fit a logistic curve to the following set of data.

Table 1. Biomass (inmm^2) at different time (in days)

Time	11	15	18	23	26	31
Biomass	.00476	.0105	.0207	.0619	.337	.74
Time	39	44	54	64	74	
Biomass	1.7	2.45	3.5	4.5	5.09	

The program, inputand the corresponding output are given below:

$n = $ Input[How many entries do you want to put]

Do[x[i]=Input["Values of x"],{i,1,n}]

Do[y[i]=Input["Values of y"],{i,1,n}]

After executing the above code in Mathematica, it is asked for the value of n and the corresponding x(time) and y (biomass). In this case, the following x and y values are given.

S={{11,0.00476},{15,0.0105},{18,0.0207},{23,0.0619},{26,0.337},{31,0.74},{39,1.7},{44,2.45},{54,3.5},{64,4.5},{74,5.09}}

ListPlot[S,PlotStyle→RGBColor[1,0,0]]

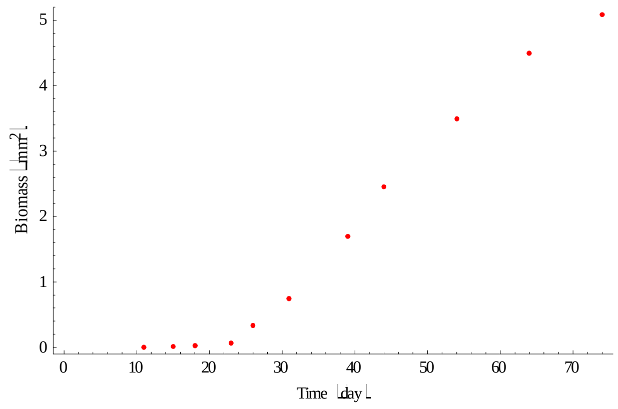

Fig. 1. Plot of the data listed in Table 1

h[t_]:=1/(1+Exp[-r*(t-t_0)])

p=Sum[y[i]*y[i],{i,1,11}]//Simplify;

q=Sum[h[x[i]]*y[i],{i,1,11}];

s=Sum[h[x[i]]*h[x[i]],{i,1,11}];

k=q/s;

Er=p-(q^2/s)

r_0=(y[2]-y[1])/(x[2]-x[1]);

r_1=(y[2]-y[1])/(x[2]-x[1]);

Do[g[j-1]=(y[j]-y[j-1])/(x[j]-x[j-1]);

If[r_0<g[j-1],r_0=g[j-1]];If[r_1>g[j-1],

r_1=g[j-1]],{j,3,11}]

r_0= 4 r_0/y[n]//N

r_1=4 r_1/y[n]//N

Plot3D[Er,{t_0,0,76},{r,r_1,r_0}]

ContourPlot[Er,{r,0,.6},{t_0,0,76},

Contours→150,Frame→True,

FrameLabel→{r,t_0}]

Simplify[k = q/s /. {r -> .1212, t_0 -> 45.79}]

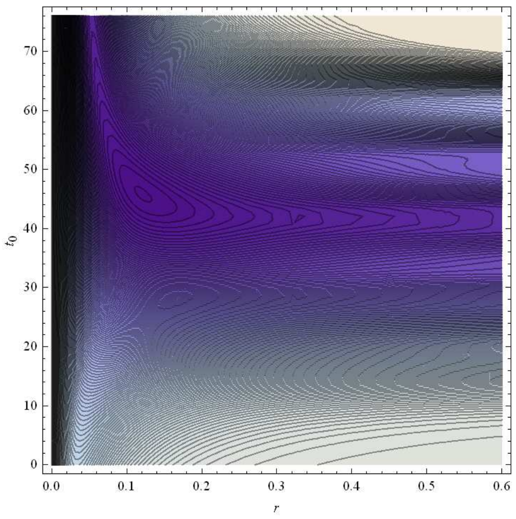

Fig. 2. Contour map of the error surface

From figure 2, we get $r = 0.1212$, $t_0 = 45.79$ by using MATHEMATICA and $k = 5.09706$ from equation (9),then

$$y(t) = \frac{5.09706}{1+257.173e^{-.1212t}} \tag{12}$$

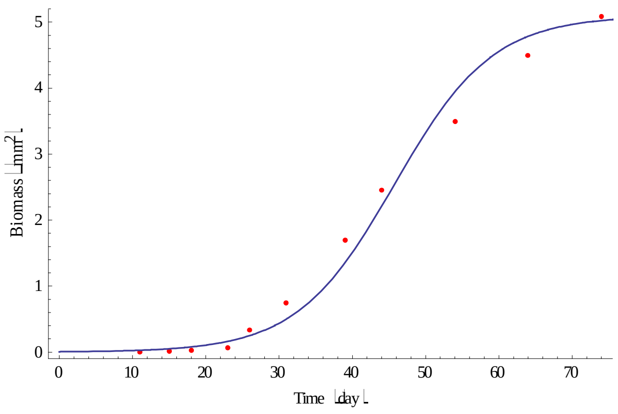

Fig. 3. Fitting of logistics curve (equation 12) and table 1.

IV. Conclusion

The method is dedicated for interpolating logistic data that exploit the intrinsic analytic properties. Accuracy of the fitting depends on the precession in choosing the co-ordinate of the minimum error, encapsulated by label curves, generated by Mathematica[7]. For the particular data we used it. It seems from figure 3 the interpolating curve fits the data better in long term prediction and not very much satisfactory initially.

References

1. Blanchard P., Robert L. Devaney, Glen R. Hall, 2002. Differential Equation, Second edition, Brooks/Cole Publishing Co.
2. Ranferi Gutie'rrez M., M.A. Reyes, and H.C. Rosu, 2010. A note on Verhulst's logistic equation and related logistics maps, J. Phys. A: Math. Theor. 43.
3. Tsoularis A., J. Wallace, 2002. Analysis of Logistics growth models, Mathematical Biosciences, Volume 179, Issue 1, 21-55.
4. Richard C., J. Fritz, 1999. Introduction to Calculus and Analysis, I, (Classics in Mathematics), John Wiley and Sons publishing Company.
5. Paul R. Holmas, 1947. Finite Dimensional Vector Spaces, Princeton University Press.
6. Wolfram S., 1991. Mathematica: A System for Doing Mathematics by Computer, Version 2.2, Addison-Wesley Publishing Company.
7. Wolfram S.,1999. The Mathematica Book,Fourth Edition, Wolfram Media/Cambridge University Press.