Application of parameter estimation in Logistic model

Jing Zhao, Gang Wu, Fan Zhang*, Xuan Zhang, Haitao Wang, Xinyu Cao
China National Institute of Standardization, Beijing, China
Corresponding author: zhaoj@cnis.ac.cn

Abstract: Firstly, Logistic model and parameter estimation are simply introduced. Secondly, we make some analysis and summary of various methods of solving Logistic model estimation, which have been put forward by scholars in recent years. These methods include: using Logistic model’s integral form, linearization, optimization method, numerical method, grey system dynamic model, population model to solve the modeling parameter identification method. What’s more, the practical applications of Logistic model are introduced such as the application in the process of urbanization.

1. Introduction
Logistic model[1] is also known as retarded growth model. It refers to that after the growth of any thing to a certain extent, the reason for the decline in the growth rate is because of the influence of some external factors such as the environment and the blocking effect on things, and its blocking effect is to increase as things grow. Suppose that

\[ h(x) = h - sx, \quad (h, s > 0), \]

Where \( h \) is the natural rate of growth of things and \( x_m \) is the maximum capacity a thing can hold, then

\[ h(x_m) = 0 \]

\[ s = \frac{h}{x_m} \]

So the natural rate of increase of things is

\[ h(x) = h \left(1 - \frac{x}{x_m}\right), \quad (h > 0) \]

The model is as follows:

\[ \frac{dx}{dt} = hx \left(1 - \frac{x}{x_m}\right) \]

And the following model

\[ \begin{cases} \frac{dx}{dt} = hx \left(1 - \frac{x}{x_m}\right), \quad (r > 0) \\ x(t_0) = x_0 \end{cases} \]

is called Logistic model.
2. Parameter estimation in Logistic model

There are some common parameter estimation methods for Logistic model. Tianjie Cao[2] gave the solution using the integral form of Logistic model. The advantage of this method is its solution process is relatively simple without repeated iterative operation. Some researchers suggested using the linearization method to solve. More specifically, [3] proposed three-point method; [4] suggested difference method; [5] given first order difference quotient method and [6] introduced three-step approach. And some other literatures proposed to using optimization method to parameter estimation methods for Logistic model. More specifically, [7] proposed Neural network identification algorithm; [8] suggested orthogonal experiment method. What’s more, [9] given numerical method; [10] introduced grey system \( GM(1,1) \) model and [11] gave population model parameter identification method. These methods have their advantages and disadvantages, for example, using the integral form of Logistic model method, its solution process is relatively simple without repeated iterative operation. The estimate error is a little larger using the linearization method.

3. Application

3.1 Logistic model and urbanization

The size of urbanization level is determined by the proportion of urban population in the national population, the higher the value is, the higher the urbanization level is. The essence of urbanization is the process from rural population to urban population. This process can be described as follows:

\[
\ln \frac{A(t)}{1 - A(t)} = \ln \frac{A_0}{1 - A_0} + Bt
\]  

(7)

where \( A(t) \) is urbanization level. \( A_0 = A(0) \), \( B > 0 \) is called the substitution rate parameter. Differentiate (7) into the following form:

\[
\frac{dA(t)}{dt} = BA(t)(1 - A(t))
\]

(8)

This Logistic model shows that urbanization speed is related to urbanization level and distance from the highest urbanization level. The urbanization process has the characteristic of accelerating development and its accelerated speed is \( B(1 - 2A(t)) \). That is, with the improvement of urbanization level, its speed is still accelerating. When accelerated speed is 0, that is \( A(t) = 50\% \), has the greatest rate of urbanization. Thereafter, due to accelerated speed is less than 0, it slows down until it reaches saturation.

Urbanization[12], refers to the advancement of the process of industrialization and social and economic development. The normal urbanization process will experience the process of urbanization, suburban urbanization, reverse urbanization and re-urbanization. After the second world war, the world urbanization has developed rapidly. After 1980, urbanization in developed countries began to slow down and was affected by informatization, presenting the phenomenon of coexistence of urbanization and reverse urbanization. At this time, the rapid growth of urbanization in developing countries has constituted the main body of urbanization.

3.2 Discussion about the saturation level of urbanization

Dr.Karmeshu [13] pointed that it is doubtful that the Logistic model has an overall population urbanization trend. According to the data in table 1, it can be found that the saturation level of urbanization is actually related to the urban demographic standard of a country. Singapore's urbanization level is 100%. The urbanization level of Canada, United States USA, United Kingdom and Australia are about 80%, 81%, 90% and 88% since the 21st century. The urbanization level of Japan is about 66%. However, the urbanization level of China and India are about 42%, 29%. From Table 1, we can conclude that the maximum value of urbanization level in some developed countries is above 85%.
When the urbanization level reaches the maximum, the population proportion of the city will fluctuate around the maximum over time. In terms of demographics, China’s urbanization can be reflected by two criteria: urban population and non-agricultural population, but the two measures represent very different levels of urbanization. If the standard is non-agricultural population, the urbanization level can only be close to, but not reached 100%. Therefore, the C value of the Logistic model can be above 85%.

Table 1 The proportion of urban population in some countries

| Country or Area | 2000  | 2005  | 2007  | 2008  | 2009  |
|-----------------|-------|-------|-------|-------|-------|
| China           | 35.8  | 40.4  | 42.2  | 43.1  | 44    |
| Japan           | 65.2  | 66.3  | 66.5  | 66.6  |       |
| Singapore       | 100   | 100   | 100   | 100   | 100   |
| Canada          | 79.5  | 80.1  | 80.3  | 80.4  | 80.5  |
| United States   | 79.1  | 80.8  | 81.4  | 81.7  | 82    |
| Italy           | 67.2  | 67.6  | 67.9  | 68.1  | 68.2  |
| United Kingdom  | 89.4  | 89.7  | 89.9  | 89.9  | 90    |
| Australia       | 87.2  | 88.2  | 88.6  | 88.7  | 88.9  |

If the urbanization saturation level is denoted as $C$, then (7) and (8) are:

$$\ln \frac{A(t)}{C - A(t)} = \ln \frac{A_0}{C - A_0} + Bt \quad (9)$$

$$\frac{dA(t)}{dt} = \frac{B}{C} A(t)(C - A(t)) \quad (10)$$

The analytic solution of the model is as follows:

$$A(t) = \frac{C}{1 + De^{-Bt}} \quad (11)$$

The parameters B, C and D in this model need to be estimated. There are two estimation methods. The first one is estimating C first then estimating B, D. The second one is estimating three parameters simultaneously. When C cannot be decided exactly, it is necessary to choose the best method to estimate C. However, if we estimate C using observation sequence, there may be the following problems: (1) When observed sequence is far away from C level the bias of estimation is bigger; (2) When observed sequence is small and random the accuracy of the C value estimation will affect the level of prediction; (3) In the prediction of urbanization level, there may be the case of C > 1. Based on the above analysis and the data presented in Table 1, we choose C = 85% and C = 100%.

When C is determined, we use two-step least square parameter to estimate B and D as follows:

Denote $\hat{A}(t)$ as the urbanization level of time t, the true value is $\hat{A}(t)$, then

$$\hat{Y}(t) = \frac{C - \hat{A}(t)}{\hat{A}(t)} \quad (12)$$

Equation (11) can be linearized as:

$$\ln \hat{Y}(t) = \ln D - Bt \quad (13)$$

First we estimate D. For two adjacent observation moments $t_k$, $t_{k+1}$, we have

$$\ln \hat{Y}(t_k) = \ln D - Bt_k, \quad k = 1, 2, \ldots, n - 1 \quad (14)$$
\[
\ln \tilde{Y}(t_{k+1}) = \ln D - Bt_{k+1}, k = 1, 2, \cdots, n - 1
\]  
(15)

\[
t_{k+1} \ln \tilde{Y}(t_k) - t_k \ln \tilde{Y}(t_{k+1}) = \ln D(t_{k+1} - t_k)
\]  
(16)

According to the least square principle, let

\[
L = \sum_{k=1}^{n-1} \left[ \left( t_{k+1} \ln Y(t_k) - t_k \ln Y(t_{k+1}) \right) - t_{k+1} \ln \tilde{Y}(t_k) - t_k \ln \tilde{Y}(t_{k+1}) \right]^2
\]

\[
= \sum_{k=1}^{n-1} \left( t_{k+1} \ln Y(t_k) - t_k \ln Y(t_{k+1}) \right)^2 - \ln D(t_{k+1} - t_k) \rightarrow \text{min}.
\]  
(17)

Let \( \frac{\partial L}{\partial \ln D} = 0, \Delta t_k = t_{k+1} - t_k \), we can get:

\[
D = \exp \left[ \frac{1}{\sum_{k=1}^{n-1} \Delta t_k \ln Y(t_k) - t_{k+1} \ln Y(t_{k+1}) - t_k \ln Y(t_k)} \right].
\]  
(18)

Then we make the estimate of B as D. According to the least square principle, let

\[
L = \sum_{k=1}^{n-1} \left[ \ln Y(t_k) - \ln \tilde{Y}(t_k) \right]^2 = \sum_{k=1}^{n-1} \left[ \ln Y(t_k) - (\ln D - Bt_k) \right]^2 \rightarrow \text{min}
\]  
(19)

Let \( \frac{\partial L}{\partial B} = 0 \), we can get:

\[
B = \frac{1}{\sum_{k=1}^{n} t_k^2} \left( \ln D \sum_{k=1}^{n} t_k - \sum_{k=1}^{n-1} t_k \ln Y(t_k) \right)
\]  
(20)

When \( t_k = k \ (k = 1, 2, \cdots, n) \), \( \Delta t_k = 1 \), (18) and (20) can be simplified into:

\[
D = \exp \left( \frac{2}{n-1} \sum_{k=1}^{n-1} \ln Y(k) - Y(n) \right)
\]  
(21)

\[
B = \frac{3n(n + 1) \ln D - 6 \sum_{k=1}^{n-1} k \ln Y(k)}{n(n + 1)(2n + 1)}.
\]  
(22)

Acknowledgements.
This research was supported by China National Institute of Standardization through the “special funds for the basic R&D undertakings by welfare research institutions” (522020Y-7475, 522019Y-6781, 522018Z-6573).

Reference
[1] Du Xiufeng, Zhang Jian, Zhang Hongmin. Mathematical Modeling [M]. Beihang University Press, 2011.
[2] Cao Tianjie. A Method for Estimating Logistic Model Parameters [J]. Journal of Taiyuan University of Technology, 1998, 29 (5):551-553.
[3] Pearl R and Reed LJ. The Logistic Curve and the Census Count [J]. Science, 1930, 72(1868):399-401.
[4] Zhang Yuanming, Ge Junyi. Estimate the Parameters in the Logistic Model [J]. Journal of Sichuan Institute of Animal Husbandry and Veterinary Medicine, 1994, 8 (2):46-52.
[5] Yang Yunqing, Miao Yaoyuan. Discussion on Fitting Method of Animal Growth Curve. Acta Veterinarian et Zootechnica Sinica, 1992, 3:219-224.

[6] Zhang Haisong, Wu Zhanwei. Studies on Populations of Nilaparvata Lugens with Logistic Curves in a Broad Sense [J]. Entomological Knowledge, 1997, 34 (4):193-195.

[7] Yu Dejiang. Neural Networks are Used for a Class of Curve Prediction [J]. Forecasting, 1992, 1:58-61.

[8] Yu Keqing, Zhao Jingzhao. The Best Fitting of a Kind of Population Ecological Curve. Natural Enemies of Insects, 1992, 1:33-39.

[9] Wu Xinyuan. A Numerical Method Fitting the Logistic Curve [J]. Journal of Biomathematics, 1990, 1:26-32.

[10] Li Liucang, Xv Wentian, Cai Xiangzhan. Discussion on Gray system GM (1, 1) model [J]. Mathematics in Practice and Theory, 1993, 1:15-22.

[11] Chen Huahao, Xv Wenke. A Method of Identifying Parameters of Population Model [J]. Journal of Biomathematics, 1990, 1:51-57.

[12] Wang Yuanfei, Zhang Chao. Logistics Model Parameter Estimation and Urbanization Level Prediction in China [J]. Economic Geography, 1997, 17(4):8-13.

[13] Karmeshu, Huang Yefang. Statistical model of Urbanization demographic [J]. Progress in Geography, 1992, 1:58-60.