JOURNAL OF
EMERGING INVESTIGATORS

Article

Population Forecasting by Population Growth Models based on MATLAB Simulation

Jingze Li[1], Hien Tran[2], Jie Li[3]

[1] The Shattuck St. Mary's School, 1000 Shumway Ave, Faribault, MN, USA
[2] The Department of Mathematics, North Carolina State University, North Carolina, USA
[3] The School of Mechatronical Engineering, Beijing Institute of Technology, Beijing, China

SUMMARY

Population forecasting is a necessary effort to understand population growth, which affects various aspects of a country's society and economy, including future demand for food, water, energy, and services. Mathematical models are commonly used to understand the interplay of the migration, birth and death rates on population growth. Mathematical models help population forecasting by capturing statistical trends from historical datasets. However, these need to be carefully compared to understand the implications of different model formulations in predicting future population, which the models have not seen or were trained on. In this work, we have compared two common population growth models, namely Malthusian law and Logistic law, using US population data from 1951 to 2019. By formulating a least-squared curve fitting problem, the birth and death rates can be estimated using MATLAB software. MATLAB simulations showed that the Logistic law of population growth yields smaller sum of squared residuals than the Malthusian law. In this case, a better population model may be beneficial in the social science, such as political science and sociology.

INTRODUCTION

In view of the world population continuously evolving in the last century, population prediction becomes more and more important in policy making, economic planning, education, and so on [1]. Population is affected by numerous factors such as policy, economy, and culture. Therefore, it is hard for demographic researchers to analyze each factor. Among all principles, historical data are the significant foundation in forecasting. By analyzing the inherent tendency inside the historical data, a comprehensive and reasonable forecast can be made without discussing each factor that affects the population. In order to achieve satisfying forecast performance, authentic historical data are needed. The Malthusian model and the Logistic model are two widely used methods for population forecast and have shown good performance. The Logistic model was firstly introduced in 1837 by the Dutch biomathematician Pierre Verhulst [2, 3]. Miranda [4] later developed the Logistic model which has proved to be well suited for population forecasting. As shown in this paper, predictions obtained via the Logistic model are in quite acceptable agreement with current official forecasts from the US Census Bureau.

The matrix laboratory (MATLAB) is primarily intended for mathematical and scientific computation and is a powerful programming software for problem solving. MATLAB allows sophisticated functions, matrix manipulations, algorithmic design, and so on to be implemented in a relatively simple manner [5]. The growth of a population is often described by ordinary differential equations (ODEs), which can be effectively solved by MATLAB. Given the importance of forecasting population and the competitive ability of MATLAB on dealing with mathematical problems encountered in real life, the goal of our research was to figure out an accurate population growth model by utilizing MATLAB. We hypothesized that the Logistic law, which considers the death rate to be a function of population size, would be more accurate and realistic than the Malthusian law in which the death rate is a constant for a long term prediction. To test this hypothesis, we obtained the US population data from 1951 to 2019 from the US Census Bureau. Then, we adopted the data for numerical verification and to compare the accuracy of the mathematical models by the Malthusian law and the Logistic law for population predictions. Furthermore, to determine whether the Logistic law gives a statistically significant improvement in the fit as compared to the Malthusian law, we conducted an F-test. In order to verify and bolster the conclusion, we applied the F-test to 20 different countries to test if the Logistic model can be widely used. The population data of the countries were obtained originally from United Nations - World Population Prospects [10].

RESULTS

We utilized the mathematical models, described by ODEs using the Malthusian law and the Logistic law, to describe the US population growth and to predict future population. The Logistic law considers the death rate as a function of population size and the Malthusian law assumes a constant death rate. We adopted the ODE45 solver in MATLAB to solve these two models, utilizing for data the US population from 1951 to 2019 [6] to determine the birth and death rates and to show how well the model fit the data.

Compared to the actual populations sizes, the Logistic law is more accurate and realistic than the Malthusian law for predicting the US population from 1951 to 2019 as sum of squared residual for the Logistic law 8.78×10^{14} is smaller than the sum of squared residual for the Malthusian law

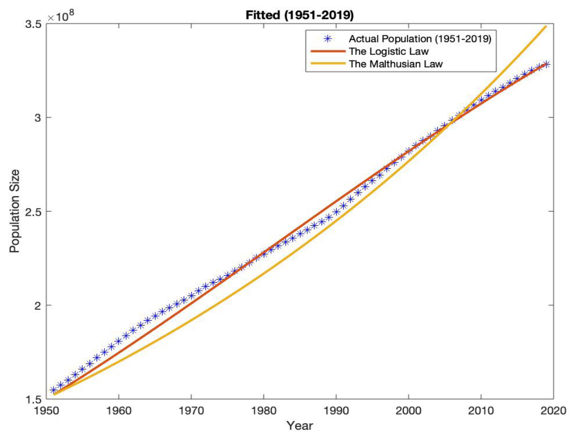

Figure 1: Fitted solutions of the mathematical models using the Malthusian law and the Logistic law using the US resident population data from 1951 to 2019. The parameters values found from *fminsearch* for the Malthusian law (r = 0.0122) and the Logistic law (b = 0.0235, n = 5.0471×10^{-11}).

6.12 x 10^{15} (**Figure 1**). In addition, near the year 2019, the predicted population by the Malthusian model continues to grow exponentially, while the predicted population by the Logistic model starts to exhibit diminishing growth. To further demonstrate that the Logistic growth model gives a statistically significant improvement in the fit, we performed an F-test. Using the degrees of freedom of 1 and 67 in the numerator and denominator, respectively, we calculated an F-statistic value of 400.22. The F-value is much larger than the F-critical value of 3.984 at 5% significance level; hence, we can reject the null hypothesis at 95% level of confidence. In other words, the Logistic model fits the actual data better than the Malthusian model does according to the F-test.

Using the birth and death rates obtained from data fitting, we obtained the predicted US population for the year 2020 from both models, which yielded 328.60 million and 348.80 million from the Logistic and the Malthusian models, respectively. Compared to the prediction value of 329.73 million from the US Census Bureau [6], the Logistic model gives a better prediction than the Malthusian model.

To further analyze the fidelity of the Logistic model, we repeated the above analysis for ten countries with high Human Development Index (HDI) and ten countries with low to medium HDI (see Method Section, "Analysis with different countries"). **Table 1** summarizes the sum of squared residuals corresponding to the model using the Malthusian law (SSR_M) and the Logistic law (SSR_L) and F-values of 20 countries. **Table 1** shows that only one country, Norway, has an F-value (2.62) lower than the critical value of 3.984. In this case, Norway's population is still in the period of accelerating growth and, hence, the Malthusian model is a good model for its population prediction (see the discussions in Section

Table 1: F-test results

	Country	SSRM	SSRL	F value	Result
Very high HDI countries	Norway	$6.67*10^{11}$	$6.42*10^{11}$	2.62	fail to reject
	Switzerland	$5.62*10^{12}$	$3.36*10^{12}$	45.15	reject
	Argentina	$1.28*10^{14}$	$2.65*10^{12}$	3169.23	reject
	Canada	$2.73*10^{14}$	$2.93*10^{13}$	557.27	reject
	United States	$6.13*10^{15}$	$8.78*10^{14}$	400.22	reject
	Australia	$5.23*10^{13}$	$1.34*10^{13}$	194.57	reject
	Iceland	$1.21*10^{10}$	$2.17*10^{9}$	306.63	reject
	Sweden	$1.83*10^{12}$	$1.27*10^{12}$	29.54	reject
	Singapore	$4.06*10^{12}$	$1.88*10^{12}$	77.69	reject
	South Korea	$1.69*10^{15}$	$2.94*10^{12}$	38446.61	reject
Low and medium HDI countries	Kenya	$1.12*10^{14}$	$9.18*10^{12}$	750.43	reject
	Myanmar	$7.86*10^{14}$	$2.68*10^{13}$	1898	reject
	Philippines	$2.10*10^{15}$	$1.48*10^{13}$	9439.76	reject
	India	$9.21*10^{16}$	$1.51*10^{16}$	341.66	reject
	Guatemala	$1.50*10^{13}$	$9.80*10^{11}$	958.51	reject
	Morocco	$3.89*10^{14}$	$7.02*10^{12}$	3645.68	reject
	Vietnam	$3.14*10^{15}$	$3.38*10^{13}$	6157.26	reject
	Guyana	$1.16*10^{12}$	$1.69*10^{10}$	4531.81	reject
	Bhutan	$1.53*10^{11}$	$1.01*10^{10}$	947.95	reject
	Bangladesh	$3.74*10^{15}$	$2.34*10^{14}$	1003.85	reject

The sum of squared residuals corresponding to the model using Malthusian law (SSR_M) and Logistic law (SSR_L) and F values of 20 countries in terms of the F-test (F-critical value at the confidence level of 95% equals to 3.984).

"Methods"). For all other nineteen countries sampled, F-values were much larger than the critical value. This result further confirms our hypothesis that, in general, the Logistic model provides a better population prediction than the Malthusian model.

DISCUSSION

Comparing the actual US population and the predicted US populations from the Malthusian law and the Logistic law of population growth, we can see that the logistic model is more in accord with the actual population than the Malthusian model is (**Figure 1**). In addition, the Logistic model also predicted a carrying capacity value of 465.61 million. Furthermore, we can visualize the simulation results that when the population size is smaller than half the carrying capacity, the population's growth trend is in an accelerated growth period. After crossing this cut-off point, the population's growth rate decreases and eventually reaches zero near the carrying capacity, which is the period of diminishing growth. This characteristic of population growth using the Logistic law, which is depicted by an S-shaped curve, seems to make sense practically. The major difference between the Malthusian law and the Logistic law is that the Logistic law takes the dynamic death rate into consideration. This death rate can be influenced by many factors. As the population size gets very large, individual members are now competing for limited resources such as food, living space, economic conditions, and medical care. Hence, population does not grow as fast as it does at the beginning, when the population size is small. One way to slow down the growth of the population for large population size is to make the death rate a function of population size as was done with the Logistic law. This means that as population grows larger, the death rate will also become larger which, consequently, will slow down population growth, so the population cannot grow at a constant or an increasing growth rate. In contrast, the model of the Malthusian law does not show the diminishing growth. We can see the estimation based on Malthusian law exceeds the actual population around 2005 and keeps growing at an exponential rate from **Figure 1**. This means that the model of the Malthusian law can only predict a very short period of population pattern.

We assumed that the actual population of US resident population from the US Census Bureau was reliable to test our predictions based on the Malthusian Law and the Logistic law. According to the US Census Bureau, the net growth rate in US population size changes yearly, from -0.06% in 1918 to 2.12% in 1910 [6]. The mean US population growth rate is 1.24%, and the median is 1.16%. Data fitting yielded a net growth rate of 1.22% for the Malthusian law model which is close to the US Census Bureau's estimation. However, the exponential growth characteristic of the Malthusian law yields gross exaggeration of population value for long term prediction into the future.

MATERIALS AND METHODS

In the literature, there are two common approaches for modeling the population size: the Malthusian law of population growth and the Logistic law of population growth. They are both theoretically sound and can be applied to practical censuses. We start with the US population in 1950 (about 152.27 million) as an initial condition and fit the two mathematical models to the US population data from 1951 to 2019.

The Malthusian Law for Population Growth

Let $P(t)$ denotes the population size at any given time t. To simplify this sophisticated problem, we assume two major impacts that would change the population, *i.e.*, the birth rate and the death rate. In the Malthusian law for population growth, the birth rate and the death rate are assumed to be proportional to the population size with the proportionality constants denoted by b and d, respectively. The change in population at any time would be described by:

$$\frac{dP(t)}{dt} = (b-d)P(t) = rP(t) \qquad (1)$$

In Equation (1), the birth rate term, $bP(t)$, is positive and will increase the population size. On the other hand, the death rate, $dP(t)$, is negative, and will decrease the size of the population. The net effect, $r = b - d$, will increase the population size if b is larger than d (more birth than death) and will decrease the population size if b is smaller than d. When $b = d$, the rate of change of the population is zero, and the population will remain constant in size.

For the initial condition, we input the US resident population in 1950. That is, t = 1950 is the initial time and $P(0) = 152.27\times10^6$.

The Logistic Law for Population Growth

The Logistic model was first developed by Pierre Verhulst [2, 3]. As discussed in this paper, for a large population it is more realistic to consider the death rate coefficient d to be a function of population size since more people will die due to limited resources, diseases, epidemics, and so on. A simple model to account for this observation is to assume the death rate coefficient to be proportional to the population size; that is, $d = nP(t)$. Hence, it is assumed when the population size is small the death rate coefficient is small and when the population size is large the death rate coefficient is large, which will slow down the population growth. Thus, the change in population size can be described as follows:

$$\frac{dP(t)}{dt} = \big(b - nP(t)\big)P(t) \qquad (2)$$

By factoring the constant b, Equation (2) can be rewritten as

$$\frac{dP(t)}{dt} = r\left(1 - \frac{P(t)}{P(m)}\right)P(t). \qquad (3)$$

Here, $r = b$ and $P(m) = b/n$. In this form and by comparison with the mathematical model for the Malthusian law of population growth (1), the net per capita growth rate, $r(1 - P(t)/P(m))$, is a function of the population size $P(t)$ instead of a constant as in Equation (1). The parameter, $P(m)$, is called the carrying capacity of the human population or, equivalently, the maximum population size. It is noted that the population will grow in size if $P(t) \leq P(m)$ (that is, the net growth rate is nonnegative). Furthermore, when the population is small relative to $P(m)$, the growth rate is large (period of accelerated growth). On the other hand, if the population size is larger and is closer to $P(m)$, the growth rate is still positive but getting smaller (period of decelerated growth). Finally, when t gets larger enough to the point that $P(t) = P(m)$ and $(1 - P(t)/P(m)) = 0$, the growth rate will be zero, and the population size is of constant size given by $P(m)$.

With the initial condition $P(0) = P_0$, the solution of Equation (3) is $P(t) = \frac{P(m)}{1+\left(\frac{P(m)}{P_0}-1\right)e^{-rt}}$ [7].

When $t \rightarrow \infty$, we obtain $e^{-tr} \rightarrow 0$ and $P(t) \rightarrow P(m)$. If we differentiate both sides of Equation (3) with respect to t, we obtain the following equation:

$$\frac{d^2P(t)}{dt^2} = r^2\left(1-\frac{P(t)}{P(m)}\right)\left(1-\frac{2P(t)}{P(m)}\right)P(t) \quad (4)$$

Equation (4) is an expression for the second derivative of $P(t)$, which will give us information on the concavity of the solution curve $P(t)$. In particular, for $P(t) < P(m)$, we observe:

1) When $P(t) < P(m)/2$, we have $d^2P(t) / dt^2 > 0$, which means the solution curve, P(t), is concave.
2) When $P(t) > P(m)/2$, we have $d^2P(t) / dt^2 < 0$, which means the solution curve, P(t), is convex.
3) When $P(t) = P(m)/2$, we have $d^2P(t) / dt^2 = 0$, which corresponds to an inflection point.

Similar to the mathematical model by the Malthusian law, we take the US population of the year 1950 as the initial condition, that is, $P(0) = 152.27\times10^6$.

Solving ODEs using MATLAB

MATLAB is a powerful mathematical tool for curve-fitting and solving ODEs. For a system of first order ODEs, an .m file that contains the descriptions of ODEs is first created. Then, by specifying input parameters (parameters in the model, *e.g.*, *r*, *b*, and *n*), initial conditions, and the time interval, where we want to compute the solution, the MATLAB's built-in ODE solver, *i.e.*, ODE45 was used to approximate the solution of the before-mentioned ODE. Besides, MATLAB was also used to visualize the solution and we used ODE45 here. Although MATLAB is powerful, it can only solve first order ODEs. In order to solve higher order ODEs, extra variables were introduced to convert a high order ODE to a system of first order ODEs [7].

Curve Fitting and F-Test

When performing the curve fitting for population growth, the *fminsearch* function in MATLAB is adopted to find the best Logistic and Malthusian growth curve (that is, the best parameters *r*, *b*, and *n*) that can fit the data. *fminsearch* is a Nelder-Mead algorithm to minimize a scalar function of several variables, which requires an initial guess of the unknown parameters. fminsearch was used to find the model parameters to give the best fitting curve via minimizing the sum of squared residuals between the data and the model solution as described by: $min_{r,b,n} J = \sum_{i=1}^{69}\left(y_i^s - y_i^{data}\right)^2$.

In this case, i denotes the year range, y_i^{data} represents the human population data at year i, and y_i^s denotes the model (Malthusian law and Logistic law) population solution at year i. When using the *fminsearch* function, the model parameter values corresponding to the period of 1951 and 1965 were selected as the initial guesses. As to the convergence criteria, we adopt the default options of the *fminsearch* function which used both the termination tolerance on the function value and on the parameter value. To be specific, the *fminsearch* function terminates when the tolerances on the function and parameter values are smaller than or equal to 10^{-4}.

In order to identify the mathematical model that best fits the population data using least squares formulation, we used an F-test. To begin, the null hypothesis, H_0, assumes that $n = 0$ (that is, the nonlinear death rate term in the model using Logistic law is not important, or, equivalently, the mathematical models using the Malthusian and Logistic laws are the same). The F-statistic is then computed from the following formula

$$F = \frac{(SSR_M - SSR_L)/m}{SSR_L/(n-m-1)},$$

where SSR_M and SSR_L are the sum of squared residuals corresponding to the model using Malthusian law and Logistic law, respectively. m is the number of restrictions (in our case, $m = 1$) and n is the number of observations (in our case, $n = 69$). It is noted that $SSR_L < SSR_M$. The question is whether the sum of squared residual associated with the model using the Logistic law is significantly less than the sum of squared residual associated with the model using the Malthusian law. The answer to this question depends on how the F-statistic value computed from the formula given above compares to the F-critical value from the F-table at the required level of significance. If the F-statistic is greater than the critical value at the required level of significance (that is, SSR_L is significantly less than SSR_M), we reject the null hypothesis.

Analysis with different countries

The countries were selected based on their HDI (Human Development Index) which is a measure that accounts for more than incomes. HDI takes many factors into account, such as education, health, and economics, to create a score between 0 and 1 [9]. A score over 0.800 is very high, a score between 0.700 and 0.799 is high, a score between 0.550 and 0.699 is medium, and a score lower than 0.550 is low. We selected ten countries with very high HDI and ten countries with medium and low HDI. Each country has a generally increasing population and available data from 1950 to 2019.

ACKNOWLEDGEMENTS

I would like to express my special thanks of gratitude to my teacher Dr. Yang who gave me the golden opportunity to spend wonderful time in North Carolina to study. Also special thanks to Guangming Wang, a physics Ph.D. student, and Yahe Yu, a mathematics Ph.D. student at NC State University for offering critical help during my projects.

Last but not the least, I would also like to thank my parents, who helped me a lot in finalizing this opportunity within the limited time.

Received: October 13, 2019
Accepted: August 22, 2020
Published: August 31, 2020

REFERENCES

1. Lassila, Jukka, et al. “Demographic forecasts and fiscal policy rules.” *International Journal of Forecasting*, vol.30, no.4, 2014, pp. 1098-1109.
2. Pierre François Verhulst. “Recherches mathématiques sur la loi d'accroissement de la population.” *Nouvelle mémoire de l'Academie Royale de Sciences et Belle-Lettres de Bruxelles [i.e. Mémoire Series 2]*, vol. 18, 1845, pp. 1–42.
3. Pierre François Verhulst. “Deuxiéme mémoire sur la loi d'accroissement de la population,” *Mémoire de l'Academie Royale des Sciences, des Lettres et de Beaux-Arts de Belgique*, vol. 20, 1847, pp. 1–32.
4. Miranda, Luiz Carlos M. and Carlos A.S. Lima. “On the Logistic Modeling and Forecasting of Evolutionary Processes: Application to Human Population Dynamics.” *Technological Forecasting and Social Change*, vol. 77, no. 5, 9 Feb. 2010, pp. 699–711.
5. Attarian, Adam. “The Official MATLAB Crash Course.” 14 Sep 2008.
6. *US Population by Year*, US Census Bureau, www.multpl.com/united-states-population/table/by-year.
7. Braun, Martin. *Differential Equations and Their Applications: An Introduction to Applied Mathematics*. Springer-Verlag, 1995.
8. Higham, Desmond J., and Nicholas J. Higham. *MATLAB Guide*. SIAM: Society for Industrial and Applied Mathematics, 2005.
9. “Human Development Reports.” *United Nations Development Programme, Human Development Reports*, hdr.undp.org/en/content/human-development-index-hdi.
10. “The Long Term Perspective on Markets.” *Macrotrends*, www.macrotrends.net/.