MODELING THE UNITED STATES POPULATION

David Arnold

College of the Redwoods
Mathematics Department

January 17, 1999

Abstract. An introduction to population modeling, from Malthus to Verhulst and the logistic model. Matlab activities using real data are supplied.

1 Malthusian Growth

In 1798, in his work *An Essay on the Principle of Population*[1], Thomas Malthus painted a pessimistic picture of the future. He argued that the geometrical growth of the human population would soon outstrip the arithmetical progression of the world's resources, leaving the world's population in dire straits.

Indeed, if we make the assumption[2] that the rate of growth of a population dP/dt is proportional to the population P, we have the result

$$\frac{dP}{dt} = aP, \tag{1}$$

where a is a constant of proportionality, often called the *intrinsic growth rate* of the population. Equation (1) is a *separable*, first order, ordinary differential equation, so we can separate the variables and integrate,

$$\frac{dP}{P} = a\,dt.$$

$$\ln|P| = at + C, \tag{2}$$

where C is a constant of integration. Exponentiating each side of equation (2) reveals

$$|P| = e^{at+C},$$

$$|P| = e^C e^{at},$$

$$P = \pm e^C e^{at}. \tag{3}$$

If we replace $\pm e^C$ in equation (3) with P_0, we arrive at the Malthusian population model

$$P = P_0 e^{at}. \tag{4}$$

[1] Currently available at `http://socserv2.socsci.mcmaster.ca/~econ/ugcm/3ll3/malthus/popu.txt`.

[2] Ecologists define the *per capita growth rate* as $\frac{1}{P}\frac{dP}{dt}$. In this case, one would argue that the per capita growth rate is constant, arriving at $\frac{1}{P}\frac{dP}{dt} = a$, which is identical to equation (1).

Typeset by $\mathcal{AMS}$-TEX

2 Exponential Growth

Consider a population obeying the Malthusian model

$$P = P_0 e^{at}. \tag{5}$$

If $a > 0$, the population will grow; if $a < 0$, the population will decline. In the first case, the population is said to exhibit *exponential growth*; in the second case, the population exhibits *exponential decay.*

For example, consider the two populations, both having an initial population of 1,000 people. The growth rate for the first population is $a = 0.02$ and the decay rate for the second population is $a = -0.02$. You can use the following Matlab code to produce the images in Figures 1 and 2.

```
close all
t=linspace(0,50);
P=1000*exp(0.02*t);
plot(t,P)
xlabel('t')
title('P=1000e^{0.02t}')
P=1000*exp(-0.02*t);
figure
plot(t,P)
xlabel('t')
title('P=1000e^{-0.02t}')
```

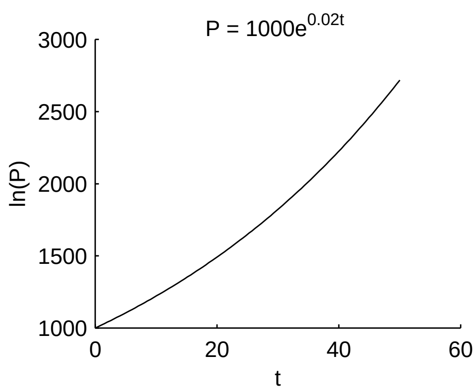

Figure 1. Exponential growth.

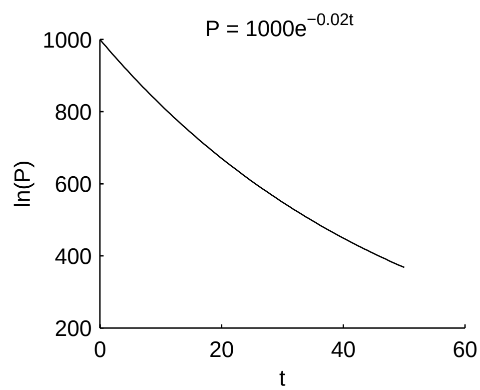

Figure 2. Exponential decay.

Suppose, for a moment, that we manipulate the model $P = P_0 e^{at}$ in the following manner.

$$\begin{aligned} P &= P_0 e^{at} \\ \ln P &= \ln\left(P_0 e^{at}\right) \\ \ln P &= \ln P_0 + \ln e^{at} \\ \ln P &= \ln P_0 + at \end{aligned} \tag{6}$$

Equation (6) indicates that if we plot the natural logarithm of the population versus the time, the graph will be a line with intercept $\ln P_0$ and slope a. The following code should produce images similar to those shown in Figures 3 and 4.

```
close all
t=linspace(0,50);
P=1000*exp(0.02*t);
plot(t,log(P))
xlabel('t')
ylabel('ln(P)')
title('ln P = ln 1000 + 0.02t')
P=1000*exp(-0.02*t);
figure
plot(t,log(P))
xlabel('t')
ylabel('ln(P)')
title('ln P = ln 1000 - 0.02t')
```

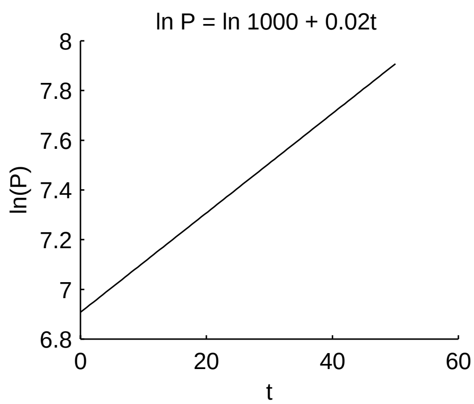

Figure 3. Exponential growth.

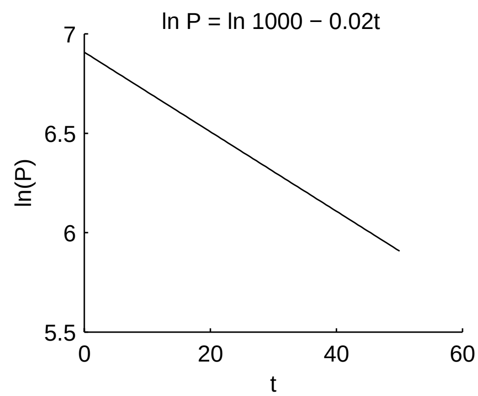

Figure 4. Exponential decay.

When a scientist suspects that her data may obey an exponential growth or decay model, she will plot her data on a sheet of semilogarithmic graph paper. If the resulting graph closely resembles a linear pattern, then the data is probably best fit with an exponential model.

Readers can examine an excellent tutorial on the use of semilogarithmic graph paper at

`http://physics.hallym.ac.kr/education/general/tutorials/GLP/`

If you don't have any semilogarithmic graph paper available, you can create your own with a piece of freeware available at

`http://perso.easynet.fr/~philimar/graphpapeng.htm`,

or you can download a single sheet of one-decade semilogarithmic graph paper at

`http://online.redwoods.cc.ca.us/instruct/darnold/DiffEq/USPopulation/onedecadesemi.pdf`.

Use the model $P = 1000e^{0.02t}$ and substitute the $t = 0, 10, \ldots, 50$ to create the following table of data.

t	P
0	1000.0
10	1221.4
20	1491.8
30	1822.1
40	2225.5
50	2718.3

Plot this data on semilogarithmic graph paper and note the linear pattern. Use the technique of the online tutorial to find the slope and intercept of this line and relate these numbers to the initial population and growth rate in the model $P = 1000e^{0.02t}$.

Matlab has a built-in capability of plotting data on semilogarithmic graph paper. For example, the commands

```
t=linspace(0,50);
P=1000*exp(0.02*t);
semilogy(t,P)
grid on
xlabel('t')
title('P=1000e^{0.02t}')
```

should produce an image similar to that in Figure 5.

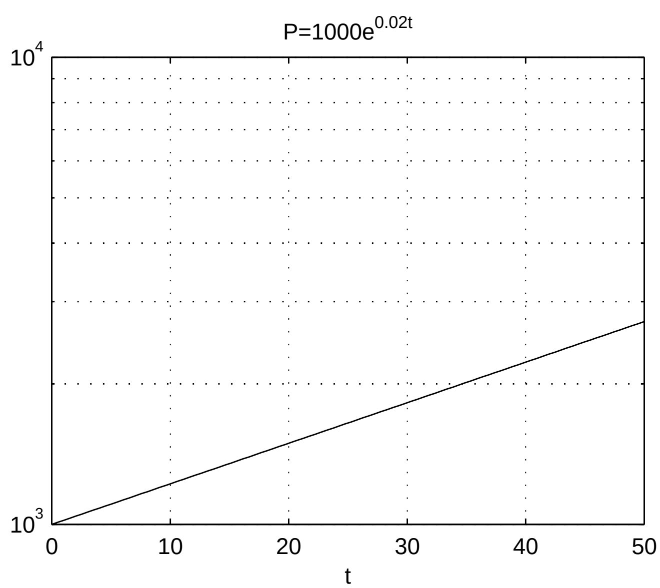

Figure 5. Matlab's `semilogy` command.

3 United States Census Data

United States census data for the years 1790 through 1980 (10 year increments, millions of people) is given in the following table.

t	Year	Pop.	t	Year	Pop.
0	1790	3.929	100	1890	62.980
10	1800	5.308	110	1900	76.212
20	1810	7.240	120	1910	92.228
30	1820	9.638	130	1920	106.020
40	1830	12.861	140	1930	123.200
50	1840	17.064	150	1940	132.160
60	1850	23.192	160	1950	151.330
70	1860	31.443	170	1960	179.320
80	1870	38.558	180	1970	203.300
90	1880	50.189	190	1980	226.550

Note that we have equated $t = 0$ with the year 1790. Enter the data for the year and population in the variables `t` and `pop`, respectively. The Matlab commands

```
t=year-1790;
plot(t,pop,'o')
xlabel('Years since 1790')
ylabel('Population (millions)')
title('US Population Data')
```

will produce an image similar to that in Figure 6.

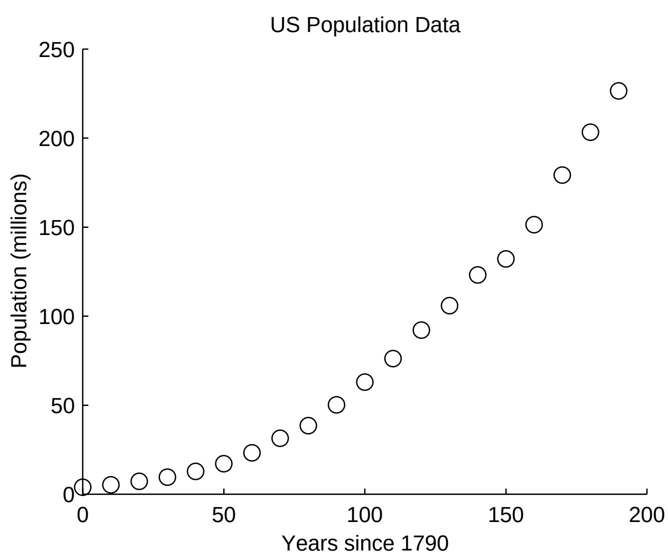

Figure 6. US census data.

Note that the data in Figure 6 appears to model exponential growth, but how can one be absolutely sure? Of course! Create a semilog plot. If the semilog plot is linear, then we know we have exponential growth. The commands

```
figure
semilogy(t,pop,'o')
xlabel('Years since 1790')
title('US Population Data')
```

will create a new figure window with a plot similar to that in Figure 7.

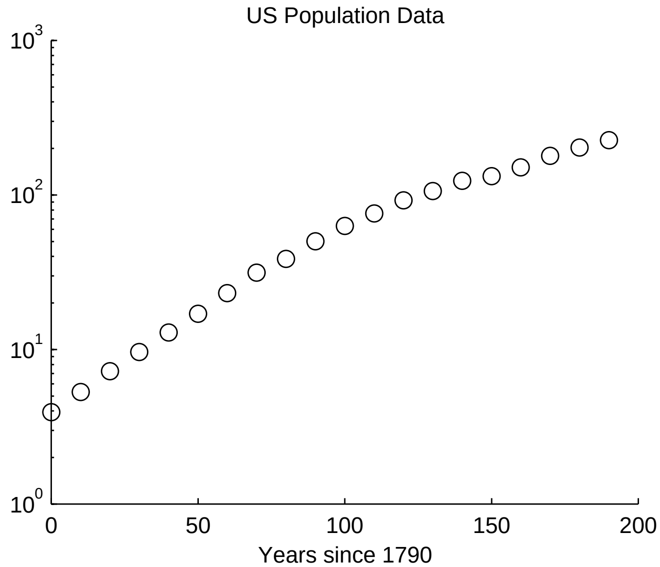

Figure 7. Semilogarithmic plot of US census data.

Alas, the plot in Figure 7 is *not* linear, so the US population cannot be accurately modeled with an exponential equation. The plot in Figure 7 was not entirely unexpected. If there is unlimited space and resources, most organisms will grow exponentially at first. However, growth will slow when food, space, and other resources become scarce.

4 Linear Least Squares

If you reexamine Figure 7, you will note that the first eight data points appear to follow a linear pattern. How can we go about finding a line the "fits" these first eight data points in an optimal manner? The answer lies in a method called *linear least squares*.

Recall equation (6), repeated here for emphasis.

$$\ln P = \ln P_0 + at \tag{6}$$

We begin by substituting the first eight data points into equation (6), as follows.

$$\begin{aligned}
\ln 3.929 &= \ln P_0 + a(0) \\
\ln 5.308 &= \ln P_0 + a(10) \\
\ln 7.240 &= \ln P_0 + a(20) \\
\ln 9.638 &= \ln P_0 + a(30) \\
\ln 12.861 &= \ln P_0 + a(40) \\
\ln 17.064 &= \ln P_0 + a(50) \\
\ln 23.192 &= \ln P_0 + a(60) \\
\ln 31.443 &= \ln P_0 + a(70)
\end{aligned} \tag{7}$$

System (7) can be written as a matrix equation as follows.

(8)
$$\begin{bmatrix} \ln 3.929 \\ \ln 5.308 \\ \ln 7.240 \\ \ln 9.638 \\ \ln 12.861 \\ \ln 17.064 \\ \ln 23.192 \\ \ln 31.443 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 10 \\ 1 & 20 \\ 1 & 30 \\ 1 & 40 \\ 1 & 50 \\ 1 & 60 \\ 1 & 70 \end{bmatrix} \begin{bmatrix} \ln P_0 \\ a \end{bmatrix}$$

System (8) is *overdetermined* (there are more equations than there are unknowns). Most overdetermined systems do not have solutions. Consequently, we will have to be satisfied with an approximate solution, a linear least squares solution.

Note that equation (8) has the form $\mathbf{b} = A\mathbf{x}$. If we divide both sides of the equation $\mathbf{b} = A\mathbf{x}$ *on the left* by the matrix A, we arrive at the solution $A\backslash\mathbf{b} = \mathbf{x}$.

The Matlab commands

```
b=log(pop(1:8))
A=[ones(8,1),t(1:8)]
```

will create the vector **b** and the matrix A, respectively. The command

```
>> x=A\b
x =
      1.3753
    0.029514
```

provides us with a least squares solution[3] of equation (8). Consequently, $\ln P_0 = 1.3753$, so $P_0 = e^{1.3753}$, or $P_0 = 3.9561$. Note that this is very close to the population census figure taken in the year 1790.

Finally, the population growth rate is $a = 0.029514$. Thus, the first eight census years obey the growth law

(9)
$$P = 3.9561e^{0.029514t}.$$

One can test the accuracy of the fit by semilog-plotting both the original census values and the population values predicted by equation (9). The commands

```
P_0=exp(x(1))
a=x(2)
P=P_0*exp(a*t)
semilogy(t,pop,'o',t,P,'+')
xlabel('Years since 1790')
title('US Population Data')
```

should produce an image similar to that in Figure 8.

[3] It may be helpful to type `help slash` and read the resulting help file, particularly the first two paragraphs.

Figure 8. Linear least squares fits data with a line.

Note that the fit is excellent for the first eight data points, but becomes progressively worse as food, space, and other resources become scarce. The commands

```
plot(t,pop,'o',t,P,'+')
xlabel('Years since 1790')
title('US Population Data')
```

show an equally good exponential fit of the first eight data points, which grows progressively worse with the shortage or resources (See Figure 9). According to Malthus, this rapid exponential growth spelled doom for mankind. He may yet be right, but that peril did not occur in the time frame predicted by Malthus.

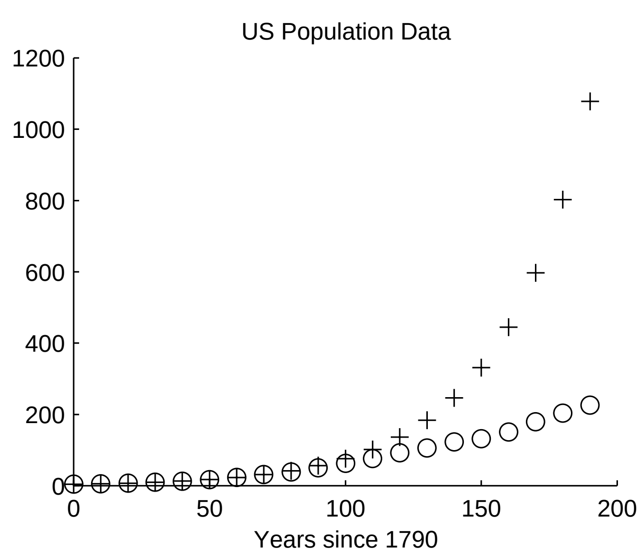

Figure 9. Original census data versus predicted data.

5 The Logistic Model

We have seen that in the presence of unlimited resources, the per capita growth rate is constant; i.e,

$$\frac{1}{P}\frac{dP}{dt} = a, \tag{10}$$

where a is a constant. However, we have also seen that as resources become scarce, the population growth tends to slow and diverge from the exponential growth predicted by Malthus. We need to adjust the model.

Perhaps the fundamental problem with equation (10) is the assumption that the per capita growth rate is constant. In fact, it might make more sense to assume that the per capita growth rate changes with time and population size.

For a first approximation, let's assume that the per capita growth rate is a function of population size alone; i.e.,

$$\frac{1}{P}\frac{dP}{dt} = f(P). \tag{11}$$

Now, what function should we choose? The simplest function that comes to mind is a linear function. Before crafting our linear function, let's make some assumptions about the behavior of our population.

(1) When the population is small, we would like the population to behave according to the Malthusian model. That is, we want the per capita growth rate to be constant, and equal to a, the intrinsic growth rate of the population.
(2) As the population increases, the per capita growth rate will decrease.
(3) As time passes, the population reaches a limiting level K, called the carrying capacity of the population environment. Populations near the carrying capacity exhibit near-zero per capita growth rate.

If we make these assumptions, then model the per capita growth rate as a linear function of P alone, the graph of f would look like that pictured in Figure 10.

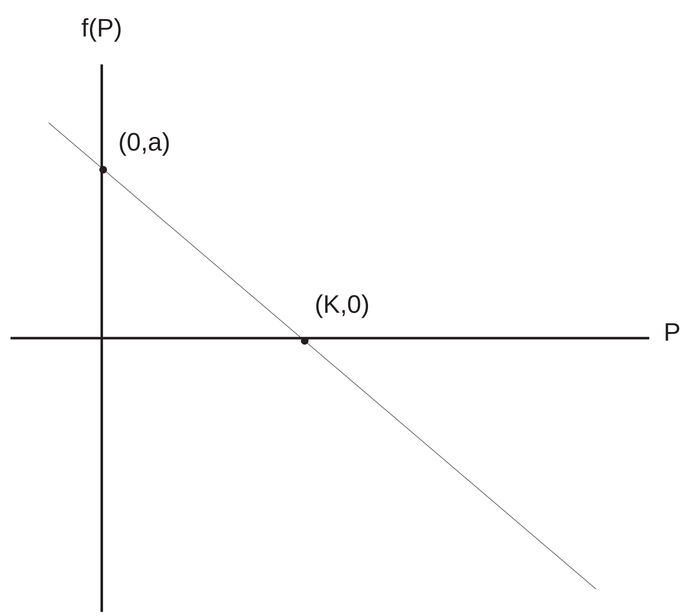

Figure 10. Per capita growth rate.

There are two important points to be made about the per capita growth rate function pictured in Figure 10.

(1) When the population is small (P is near zero), the per capita growth rate is near a. In fact, when $P = 0$, the per capita growth rate equals a.
(2) When the population is near the carrying capacity K, the per capita growth rate is near zero. In fact, when $P = K$, the per capita growth rate equals zero.
(3) If the population is less than the carrying capacity K, then the per capita growth rate is positive and the population increases. If the population is greater than the carrying capacity K, then the per capita growth rate is negative and population must decrease. This makes K a *stable* equilibrium point.

Finally, it is easy to see that the equation of the linear function shown in Figure 10 is

$$f(P) = a - \frac{a}{K}P,$$

or

$$f(P) = a\left(1 - \frac{P}{K}\right). \tag{12}$$

Finally, if we replace the per capita growth rate in equation (11) with the result found in equation (12), we arrive at

$$\frac{1}{P}\frac{dP}{dt} = a\left(1 - \frac{P}{K}\right),$$

or

$$\frac{dP}{dt} = a\left(1 - \frac{P}{K}\right)P. \tag{13}$$

In 1845, the French mathematician Pierre-François Verhulst, in his paper *Recherches mathématiques sur la loi d'accroissement de la population*, was the first to coin the phrase 'logistique' in describing equation (13), which today is commonly called the *logistic equation*.

Fitting the Logistic Equation to Data. In 1920, in *On the Rate of Growth of the Population of the United States Since 1790 and Its Mathematical Representation*, a paper presented to the *Proceedings of the National Academy of Sciences*, Pearl and Reed fit a logistic model to the United States census data given in our earlier examples population growth. In their paper, they fit the US population to the logistic equation by using three census dates to determine unknown constants. The work of Pearl and Reed was very controversial, and an fascinating history of the logistic equation in mathematical ecology is described in Sharon Kingsland's *Modeling Nature, Episodes in the History of Population Ecology.*

Gause, in his text *The Struggle for Existence*, performed a series of experiments with yeast bacteria and showed that their growth followed a pattern predicted by the logistic equation. The logistic equation is non-linear, so fitting a logistic to data is not possible using linear least squares. In the Appendix of his book (available at the Humboldt State University Library), Gause describes the procedure he used to fit a logistic to the data acquired in his experiments. The procedure involves an assumption about the carrying capacity, followed by a logarithmic transformation, making the data amenable to a linear least squares fit.

Fabio Cavallini describes a more modern approach to fitting a logistic in his paper *Fitting a Logistic Curve to Data*, presented to the *College Mathematics Journal* in 1993. It involves the use of Mathematica to minimize a function of several variables, so some experience with multivariable functions would be helpful when reading his paper. It is not difficult to translate the Mathematica commands to Matlab commands and get results similar to those found in the paper. Cavallini's method can be used to fit a wide variety of functions to data. It is a valuable method to learn.

Copies of the papers mentioned above are available in my office.