Logistic Curves: construction, fit and unicity

Valerio FRANCHINA and Roberto VACCA

(Dated: 14-16 June 1989)

International Conference on DIFFUSION OF TECHNOLOGIES AND SOCIAL BEHAVIOR, at IIASA (International Institute for Applied System Analysis), Laxenburg, Austria

I. DEFINITION OF THE PROBLEM

The logistic curve:

$$y = \frac{a}{(1 + e^{cx+B})}$$

is often used to study processes of growth of populations of biological entities and of human artefacts. Many experimental observations confirm the existence of laws aptly described by this curve. It is important, therefore, to determine the parameters of a logistic from an empiric time series, in order to be able to forecast the future of the growth process considered. In what follows the logistic will be analysed in a slightly different form:

$$y = \frac{a}{(1 + be^{cx})}$$

obtained by means of the substitution:

$$b = e^B.$$

We frequently use also the inverse form:

$$Y = cx + ln(b)$$

where: $Y = ln\left(\frac{a}{y-1}\right)$, or $Y = ln\left(\frac{a-y}{y}\right)$.

If we consider the values y/a as percentages of the asymptote a, and we define $F = 1 - y/a$, the Y take the form: $ln(F/(1-F))$, which is termed the Fisher-Pry transform. In the x, Y plane the inverse form is a straight line, whose angular coefficient is c, and intersecting the ordinates axis at $ln(b)$. The a parameter does not appear explicitly in the formula, because it is implicit in the Y values. It is commonly said that using the inverse form is equivalent to representing the logistic curve in a semilogarithmic plane, transforming it in a straight line. This is not strictly true, as any change in parameter a, entails changing the scale of the Y ordinates and hence also the metrics of the x, Y plane. So two logistic curves plotted for different values of a, are not immediately comparable in the inverse representation since they are drawn on two planes which do not coincide metrically. In the case of the direct (exponential) definition, the logistic curve has the following remarkable features:

- horizontal asymptotes, having ordinates $y = 0$ and $y = a$. The latter is commonly termed the asymptote of the logistic.

- intersection with the ordinate axis ($x = 0$) at $y = a/(1 + b)$;

- coordinates of the flex point, which is also an anti-symmetry center, given by $(-ln(b)/c;\ a/2)$.

At the flex point the logistic has its maximum slope: $y'_{max} = -ac/4$. The main effect of parameters b and c on the shape of the curve concerns respectively the position with respect to the ordinates axis (parameter b) and the slope of the central portion (parameter c). In most practical cases the abscissa is time, and the curve grows at a considerable distance from the origin. The values of parameter b are, then, rather high, so that the intersect with the ordinates axis is quite small compared to a. The position of the flex point, then, is affected by the parameter b through a logarithm, so that, even if b changes considerably, the effects on the asymptote value are quite moderate.

II. EXPERIMENTAL DETERMINATION OF THE LOGISTIC EQUATION PARAMETERS

Let us suppose we have two correlated N components vectors x_i, y_i, the y_i's being experimental measurements of the size of a population at times x_i's. We want to determine the parameters a,b,c of the logistic which fits best the data contained in the two vectors. It is well known that the problem is generally solved by applying a regression technique which, by minimising a fit indicator, permits to determine a set of parameters adequately near to those on which depends the growth law described by the logistic. In this particular context, the importance of the physical meaning of parameters a, b and c is not the same since essentially the goal of the investigation is a—the asymptote value. The non linearity of both the direct and the inverse form of the equation oblige us in any case to use more or less sophisticated iterative algorithms, which almost invariably present some snags. As a consequence many methods and recipes have proliferated often producing quite diverging estimates of the asymptote for a given series of experimental observations. This indicates that the problem is really much harder to define than we would have tended to believe, especially if we take into account the presence of inevitable errors in the measurements.

A. Method of the direct estimate of the asymptote

If we assume we know the asymptote a, then we know also the Y values for each experimental y value in the

logarithmic form of the logistic, as for each i:

$$Y_i = ln\left(\frac{a}{y_i - 1}\right)$$

This suggests an iterative method based on the following algorithm:

1. choose an initial value of a. This can be done by taking the maximum of the y_i's and multiplying it by an adequate coefficient greater than 1 (e.g., 1.2). Based on the position of the swarm of measured points, it is also possible to choose an asymptote value likely to be nearer to the real one and consequently to reduce the number of steps for convergence.

2. compute the vector of the Y_i's associated to the y_i's. Here we must watch out. It may happen that a tentative asymptote value may be smaller than one of the y_i's in the vector. In this case the method fails because the corresponding Y_i is defined as the logarithm of a negative number. (The fact that a value appears to be greater than the asymptote is a paradox. It is easy to explain, though, as due to the presence of experimental errors causing one or more y_i's to be unduly large and numerically greater than a.) apply a linear regression to the straight line $Y = ln(b) + cx$, using the swarm of N points x_i, Y_i (see Note 0.);

3. compute the normalised root mean square deviation on the set of deviations $s_i = Y_i - ln(b) - cx_i$, change the value of the asymptote a until the RMS deviation is smaller than a preset value or until it is situated in the vicinity of the absolute minimum, following an appropriate strategy in the choice of corrections made on the value of a.

This method presents two drawbacks, closely interrelated. The first concerns the scarce sensitivity of the RMS deviation to variations in the value of the asymptote. Consequently—and this is the second drawback—any procedure devised to steer the asymptote variations on the basis of the RMS deviation used as a fit indicator, does not guarantee a fast convergence. As an alternative we can use the RMS deviation of the direct form, which presents a more evident minimum when the value of the asymptote a is varied. This choice, though, is formally less correct. It is equivalent, in fact, to an attempt to look in the inverse form for that value of a, which minimises the RMS deviation in the direct form of the representation.

The following example demonstrates a case in point. Here the logistic has been generated analytically, with $a = 10$, $b = 5$ and $c = -1$.

The vectors x_i and y_i have been computed in the lower half of the logistic, without reaching the flex point, the abscissa of which is 1.609, as shown in the following table:

i	x_i	y_i
1	-8	0.0006708802
2	-7	0.0018234314
3	-6	0.0049550479
4	-5	0.0134577585
5	-4	0.0364975825
6	-3	0.0985924113
7	-2	0.2635373851
8	-1	0.6853347680
9	0	1.6666666667
10	1	3.5218742835

We have computed then the values of parameters b and c starting with $a = y_{max} * 1.01$, and proceeding to larger values of a.

In the following table we give the values of the parameters a, b and c, and those of the normalised RMS deviations computed on the direct form (dev1) and on the inverse form (dev2).

a	b	c	dev_1	dev_2
3.557	0.453	-1.252	0.182	0.088
6.000	2.504	-1.032	0.057	0.007
7.000	3.140	-1.019	0.039	0.004
8.000	3.765	-1.011	0.024	0.002
9.000	4.384	-1.005	0.011	0.001
11.000	5.613	-0.996	0.009	0.001
12.000	6.225	-0.993	0.017	0.001
15.000	8.056	-0.987	0.036	0.002
20.000	11.096	-0.982	0.057	0.003
30.000	17.166	-0.976	0.079	0.003
40.000	23.230	-0.973	0.091	0.003
50.000	29.292	-0.972	0.098	0.003
60.000	35.353	-0.971	0.103	0.003
70.000	41.413	-0.970	0.107	0.003
100.000	59.593	-0.969	0.114	0.003

Notice the scarce effect of the variations of a on dev_2, which is consequently hardly usable to steer the iterative procedure for finding the logistic's parameters. On the other hand, dev_1 (the direct form deviation) presents an acceptable sensitivity: it also tends to become constant but for much higher values of a. If we go on increasing the value of a, we get the following results:

a	b	c	dev_1	dev_2
500.000	301.961	-0.967	0.126	0.003
1000.000	604.917	-0.966	0.128	0.003
10000.000	6058.106	-0.966	0.130	0.002
100000.000	60589.991	-0.966	0.130	0.002

indicating that also dev_1 tends to a constant value (0.130).

One could object that there is no real reason for perplexity, since there is a huge discrepancy between reasonable values of a and values for which the normalised deviation becomes constant. This is certainly true for semiautomatic procedures, or when a human intervention is acceptable. But if we look for a completely automatic procedure—possibly fool-proof—the lack of sensitivity in

the behaviour of the RMS deviation represents a serious snag. Even if we decide to start from very low values of a and to try to reach the first minimum (which presumably is the correct one), we cannot say anything concerning the step to be used in this procedure. We also cannot be sure that we will not go past the point of progressive diminution of the normalised RMS deviation. A program robust enough to guarantee a correct behaviour at least in the great majority of cases, would then become quite complex and involved. Finally we should consider the fact that the behaviour of the regression algorithm applied to analytic data, is certainly more clear cut and univocal than would be the case when the quality of data is downgraded by inevitable experimental errors. We have, then, carried out on the above data a simple experiment, consisting in rounding off the values of the y_i's to the third decimal place. This corresponds roughly to assume the presence of errors of the order of .00005 of the asymptote value (which is 10 in the example), although in this case the errors obviously are not random. In any case, even with these small variations, computing parameter a, based on the rounded off data, leads to the results of the following table :

a	b	c	dev_1	dev_2
3.557	0.473	-1.227	0.178	0.093
6.000	2.622	-1.007	0.073	0.012
7.000	3.288	-0.994	0.058	0.009
8.000	3.943	-0.985	0.046	0.008
9.000	4.591	-0.979	0.036	0.007
10.000	5.236	-0.975	0.028	0.006
11.000	5.879	-0.971	0.021	0.006
12.000	6.520	-0.968	0.016	0.006
15.000	8.436	-0.962	0.014	0.005
20.000	11.621	-0.956	0.026	0.005
30.000	17.977	-0.951	0.043	0.005
40.000	24.327	-0.948	0.053	0.004
50.000	30.676	-0.947	0.059	0.004
60.000	37.023	-0.945	0.063	0.004
70.000	43.369	-0.945	0.066	0.004
100.000	62.408	-0.943	0.071	0.004
500.000	316.223	-0.941	0.081	0.003
1000.000	633.486	-0.941	0.083	0.003
10000.000	6344.221	-0.941	0.084	0.003
100000.000	63451.564	-0.941	0.084	0.002

We notice then that:

1. the minimum for dev_1 is to be found near the value $a = 15$, whereas the minimum for dev_2 is not present in the above range of a's, which reaches values of the asymptote 10,000 times larger than the theoretical value

2. the values of parameters b e c computed in the two cases are quite divergent (remember that c is an exponent)

3. the series of values used for extrapolation is adequate, since it contains values up to $x = 1$ (the flex point has abscissa $x = 1.609$) and it is formed by 10 points

4. the inverse form deviation, which from a theoretical viewpoint would have to be preferred for a correct application of the regression method, is actually useless, because it does not present a minimum.

The determination of the asymptote value is, then, extremely sensitive to the introduction of very small variations in the coordinates of empiric points. This suggests the conclusion that the experimental determination of a logistic parameters is not possible, since it is very unlikely that we can obtain empiric measurements affected by errors as small as those artificially introduced in the example above. Actually the marked difference between the asymptote determined on the basis of the artificially erroneous values and its theoretical value is due to the fact that rounding off to the third decimal figure introduces very large errors in the ordinates of the first few points (these are of the same order of magnitude of the ordinates, e.g. for the first point). We have to conclude that also the first points quite distant from the flex point give a significant contribution to the definition of the curve, as a consequence of the non linearity of the mathematical law it represents. We have introduced, then, in the y_i values, errors of the order of .005 of each ordinate (NOT of the asymptote value) rounding them off to the third significant figure—and we have obtained both for parameters and for deviations, results different from those of the theoretical case by a few tenths of one percent. This leads to the conclusion that using the direct form RMS normalised deviation may permit to obtain reasonable results, but that we should try hard to define a different fitness indicator, amenable to be used in automatic applications.

B. Analytic determination of the logistic through three given points

The logistic equation is defined by three parameters and has three degrees of freedom. It is, then, completely determined if we impose that the curve pass through three given points. In fact, if in the equation previously introduced

$$y = \frac{a}{(1 + be^{cx})}$$

we substitute $H = e^c$, we obtain:

$$y = \frac{a}{(1 + bH^x)}$$

The passage through three points with coordinates x_i, y_i, $(i = 1, 2, 3)$ leads to the three equations system:

$$y_1(1 + bH^{x_1}) = a$$
$$y_2(1 + bH^{x_2}) = a$$
$$y_3(1 + bH^{x_3}) = a$$

Eliminating parameter a between the first and second, and the first and third equation, we get the equivalent system:

$$y_1(1 + bH^{x_1}) = y_2(1 + bH^{x_2})$$
$$y_1(1 + bH^{x_1}) = y_3(1 + bH^{x_3})$$

and eliminating also parameter b we get :

$$(y_1 - y_2)y_3H^{x_3} + (y_3 - y_1)y_2H^{x_2} + (y_2 - y_3)y_1H^{x_1} = 0$$

This equation is to be solved excluding negative values of H and also the case $H = 1$, because it would annul parameter c. If the three points are equally spaced, that is $\triangle_x = x_2 - x_1 = x_3 - x_2$, we obtain an equation of second degree in $H^{\triangle_x}$, which has the solutions: 1 (unacceptable, as indicated above), and:

$$H^{\triangle_x} = \frac{y_1(y_2 - y_3)}{y_3(y_1 - y_2)}$$

from which the values of parameters a, b and c can be deduced immediately. In fact, $\triangle_x$ being known, we compute $H^{\triangle_x}$, and H is given by:

$$H = exp\Big(ln\frac{y_1(y_2 - y_3)}{y_3(y_1 - y_2)} \,/\, \triangle_x\Big)$$

We have then $c = ln(H)$. The value of parameter a is obtained using one of the relationships of the type:

$$a = (1 - H^{\triangle_x})/\Big(\frac{1}{y_3} - \frac{H^{\triangle_x}}{y_2}\Big)$$

and the value of parameter b is found in a similar way. Here care must be taken to use ratios and to avoid the direct computation of powers of the type H^x. These, in fact, could reach very high values depending on the metric scale used for the time variable x, and could create obvious problems. When we are dealing with empiric data, not all triplets of passage points are equally reliable. Moreover, equal spacing of the points along the x axis is one of the keystones of the procedure, so that errors in time measurements may also have considerable effects. These considerations have led to the formulation of the triplets method presented in the next section.

C. The triplets method

The fact that we can determine a logistic equation analytically without using iterative methods, with the only constraint of equal spacing of the three passage points, suggests an unconventional procedure for determining the three parameters. The algorithm is the following:

1. form all possible equally spaced triplets using the empiric data;

2. compute the parameters of the logistic defined by each triplet;

3. compute mean and normalised RMS deviation of the asymptotes;

4. compute the standard deviation (.) of the asymptotes;

5. eliminate the values which differ from the mean by more than 3, and repeat the procedure, if necessary, until no more values have to be eliminated.

The elimination steered by means of the value of cannot be applied to the values of the asymptotes proper, though, because their distribution is not symmetric. Consequently the only asymptotes to be eliminated would be the extremely large ones. Given the logarithmic character of the relationship, it makes sense to apply the method of progressive elimination to the set of logarithms of asymptotes. In this way it may be possible to eliminate the effects of empiric data affected by excessive errors. This algorithm may be the target of serious methodological criticism. (In particular, the elimination of some values presumably affected by excessive errors may hardly have significance in the case of a socio-economic situation, as we can imagine that sudden perturbations may deeply alter its mechanisms). However, its application to certain experimental growth phenomena has shown good accord with the values of parameters determined by means of iterative procedures. From a set of 64 empiric values (given in Note 1), we have extracted 806 equally spaced triplets, from each of which a value of the asymptote has been computed. The a values ranged from 10 to over 2 million; their mean value was 7226. After repeating 3 times the elimination of 21 asymptotes the logarithm of which differs by more than 3 from the mean value, the number of triplets was reduced to 785, the asymptotes mean value was 2988, the standard deviation was 1.28 (while it had been 1.48 before the elimination). Using other methods we had determined a value of the asymptote quite near to 3000. This indicates that the triplets method may actually supply a satisfactory solution, especially when the data include a large quantity of equally spaced observations. In these cases the virtual lack of iterations (save for the one in the elimination phase) makes the method especially fast and efficient, even if it is applied on small computers. The fact that for each equally spaced triplet an analytic relationship is available, which permits to compute the coefficients a, b and c of the logistic passing through the 3 points, suggests further refinements. A first observation concerns the value of $H^{\triangle_x}$ computed from the 3 quantities y_1, y_2 and y_3. The expression in which they appear has to be positive, so that a real and positive value of H can be obtained. This imposes a constraint, on the basis of which it is not always possible for a logistic to pass through three arbitrary points. This leads us to believe that, according to the actual position of the 3 points, the logistic's definition may be more or less credible—as is always the case when a limiting condition may be satisfied too marginally. In fact it is possible to define an indicator, termed D_{factor}, which supplies an estimate of how

reliable is the determination of the coefficients of the logistic associated to the equally spaced triplet considered. The D_{factor} can be computed in a number of equivalent ways. The basic idea is to try and evaluate the angle at which the parabola which determines $H^{\triangle x}$ in the plane $(D, H^{\triangle x})$, that is:

$$(y_1 - y_2)y_3[H^{\triangle x}]^2 + (y_3 - y_1)y_2 H^{\triangle x} + (y_2 - y_3)y_1 = D$$

intersects the D axis, at which point H assumes the sought for value. The value of the tangent of that angle (which is the parabola's derivative at the point in question), is:

$$D_{factor} = y_1(y_3 - y_2) - y_3(y_2 - y_1)$$

This value is negative, since it corresponds to the lesser parabola intersect with the D axis (the other one is always 1), and in subsequent computations this sign has to be taken into account. It is not possible to define the ranges in which the D_{factor} can be considered acceptable, as the criterion depends on the scale factors used for abscissae and ordinates. In any case, computing the D_{factor} for each triplet and examining its distribution, in general we can deduce clearly enough which triplets present an intersection angle too low with respect to the mean, supplying therefore unreliable values of the asymptote. In the example given in Note 1., the D_{factor} varies from 2 to more than 700,000, but, after elimination, the range is limited to values below 10,000, which confirms the validity of this type of indicator. If we work with triplets which are not equally spaced, the resolving curve for $H^{\triangle x}$ is not a parabola any more, but the D_{factor} continues to have the meaning we have described.

D. Variational method

The theoretical basis of this method are well known. Here we will try to apply it to the determination of the logistic's direct form. Applying the RMS minimisation to said form, the variational equations (needed to evaluate the coefficients a, b and c from a table of associated values (y_i, x_i) are very complicated. We could try to avoid the obstacle by manipulating the expression of the generic deviation S_i, until we obtain it in the form

$$S_i = a - y_i(1 + be^{cx_i})$$

In this way, though, the deviations to be minimised are multiplied by $(1 + be^{cx_i})$, which makes questionable the whole procedure. In fact it can be verified directly that in this case the RMS minimisation is not correct. On the other hand, even working with deviations obtained from the inverse form of the logistic equation, we cannot obtain neat indications apt to localise the minimum. We are obliged, then, to compute the derivatives from the direct (exponential) form. Let, then:

$$z_i = 1/(1 + be^{cx_i})$$

The derivatives of this expression with respect to the 3 unknown parameters a, b and c are :

$$0$$
$$-z_i^2 e^{cx_i}$$
$$-bx_i z_i^2 e^{cx_i}$$

We can build, now, the minimising system composed by the equations:

$$Sy_i z_i - aSz_i^2 = R_1$$
$$Sy_i z_i^2 e^{cx_i} - aSz_i^3 e^{cx_i} = R_2$$
$$Sx_i y_i z_i^2 e^{cxi} - aSx_i z_i^3 e^{cx_i} = R_3$$

where R_1, R_2, R_3 are the residues which ideally should be annulled to obtain the minimum. Carrying out the total differentiation and the subsequent passage to finite differences in the indicated system of equations, we obtain the resolving linear system (not given here due to typographic difficulties). At this point, beginning with arbitrary initial values of the coefficients a, b and c, we apply the usual iterative algorithm. The initial estimate of the three coefficients is quite important. If we choose them in an inappropriate way, the method may converge very slowly—or not converge at all. It is advisable, then, to carry out an approximate predetermination of the three values, keeping in mind their geometric significance. Coefficient a is the logistic's asymptote. It must be greater than the maximum y_i, of which it has the physical dimensions. A simple visual inspection of the empiric values may suggest a sensible value for a, greatly simplifying the procedure. The choice of b and c is less immediate. Once a has been assigned, though, it is possible to compute them passing to logarithms in the fundamental logistic and applying the RMS minimisation method given in Note 0. In the program for implementing the variational procedure, there is a gain coefficient which has been varied between .1 and .5, always obtaining a good convergence without any significant stability problems. We have not built into the program an auto-adaptive determination of the gain coefficient, since the overall performance was quite acceptable. Using the program on the data given in Note 1., the asymptote value found was 2910, quite near to those found with the other methods presented above.

III. NOTE 0. LINEAR REGRESSION FOR AN INVERSE LOGISTIC

The generic deviation has the form: $s_i = Y_i - ln(b) - cx_i$. If we annul the derivatives of the sum of squares of deviations with respect to parameters $ln(b)$ and c, we find the linear system of the second order which has the coefficients matrix:

$$\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & N \end{bmatrix}$$

The vector of known terms is:

$$\begin{bmatrix} \sum x_i * Y_i \\ \sum Y_i \end{bmatrix}$$

using the same notation. The normalised RMS deviation is defined as the square root of the quantity $\sum(s_i^2)$, divided, then, by $\sum Y_i$.

IV. NOTE 1. EXAMPLE OF APPLICATION

The data vector represents numbers of extant AIDS patients in Italy, counted at the end of each month starting in December 1983 until March 1989

5 6 8 9 12 10 10 12 13 14 12 14 18 29 38 47 52 54 58 67 66 85 89 103 108 122 142 146 160 173 192 207 227 237 262 283 289 328 357 385 424 455 490 533 560 606 650 715 757 818 872 948 1013 1108 1182 1261 1404 1482 1560 1637 1689 1765 1828 1881

Intermediate results obtained with the triplets method: Mean value of asymptote a in the 806 triplets: 7226
Values during the progressive elimination procedure:

- = 1.477 — average $ln(a)$ is: 7.239 — working on: 806 triplets

- = 1.338 — average $ln(a)$ is: 7.252 — working on: 793 triplets

- = 1.292 — average $ln(a)$ is: 7.263 — working on: 787 triplets

- = 1.278 — average $ln(a)$ is: 7.273 — working on: 785 triplets

estimated asymptote for given data set is: 2988.
Values of the asymptote before eliminations between 10 and 2,361,831 — Mean value of D_{factor} 148,817
After elimination, the values of the asymptote varied between 32 and 64,798.