4. Logistic Curve Fitting

Background for the Logistic Curve Fitting.

Fit the curve $y = f_1[x] = \dfrac{L}{1 + c\,e^{a x}}$ to the data points $(x_1, y_1),\ (x_1, y_1),\ \ldots,\ (x_n, y_n)$.

Rearrange the terms $\dfrac{L}{y} - 1 = c\,e^{a x}$. Then take the logarithm of both sides:

$$\ln\left(\frac{L}{y} - 1\right) = \ln\left(c\,e^{a x}\right) = \ln(c) + a x.$$

Introduce the change of variables: $X = x$ and $Y = \ln\left(\dfrac{L}{y} - 1\right)$. The previous equation becomes

$Y = \ln(c) + a X$ which is now "linearized."

Use this change of variables on the data points $X_k = x_k$ and $Y_k = \ln\left(\dfrac{L}{y_k} - 1\right)$, i.e. same abscissa's but transformed ordinates.

Now you have transformed data points: $(X_1, Y_1),\ (X_1, Y_1),\ \ldots,\ (X_n, Y_n)$.

Use the "Fit" procedure get Y = A X + B, which must match the form $Y = \ln(c) + a X$, hence we must have $c = e^B$ and a = A.

Remark. For the method of "data linearization" we must know the constant L in advance. Since L is the "limiting population" for the "S" shaped logistic curve, a value of L that is appropriate to the problem at hand can usually be obtained by guessing.

Example 1. Use the method of "data linearization" to find the logistic curve that fits the data for the population of the U.S. for the years 1900-1990. Fit the curve $y = f_1[x] = \dfrac{L}{1 + c\,e^{a x}}$ to the census data for the population of the U.S. Estimate the population in 2000.

Date	Populatlion
1900 July	76094000
1910 July	92407000
1920 July	106461000
1930 July	123076741
1940 July	132122446

1950	July	152271417
1960	July	180671158
1970	July	205052174
1980	July	227224681
1990	July	249464396

Solution 1.

Example 1. Use the method of "data linearization" to find the logistic curve that fits the data for the population of the U.S. for the years 1900-1990. Fit the curve $y = f_1[x] = \frac{L}{1 + c\, e^{a x}}$ to the census data for the population of the U.S. Estimate the population in 2000.

Date	Populatlion
1900 July	76094000
1910 July	92407000
1920 July	106461000
1930 July	123076741
1940 July	132122446
1950 July	152271417
1960 July	180671158
1970 July	205052174
1980 July	227224681
1990 July	249464396

Solution 1.

A limiting population L, or "carrying capacity" must be estimated. For this data the number L is not too sensitive, but must be larger than the largest ordinate so that the values $\ln\left(\frac{L}{y_k} - 1\right)$ are not complex numbers. For illustration, we choose L = 800 million.

L = 800

800

Do a series of intermediate computations.

$\{x_k\} = \{1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990\}$

$\{y_k\} = \{76.094, 92.407, 106.461, 123.077, 132.122, 152.271, 180.671, 205.052, 227.225, 249.464\}$

$X_k = x_k - 1900$

$L = 800$

$Y_k = \text{Log}[\frac{L}{y_k} - 1]$

$\{X_k\} = \{0, 10, 20, 30, 40, 50, 60, 70, 80, 90\}$

$\{Y_k\} = \{2.25269, 2.03567, 1.87403, 1.70475, 1.62038, 1.44781, 1.23196, 1.06521, 0.924554, 0.791575\}$

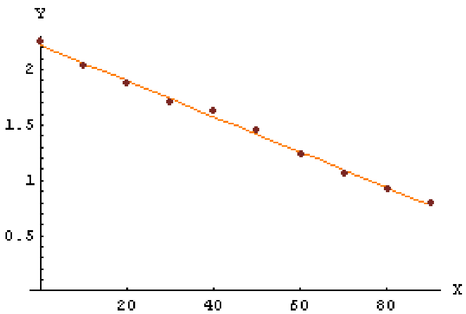
Now glue together the transformed parts to form the pairs $\{(X_k, Y_k)\}$.

$\{(X_k, Y_k)\} =$

$\{\{0, 2.25269\}, \{10, 2.03567\}, \{20, 1.87403\}, \{30, 1.70475\}, \{40, 1.62038\}, \{50, 1.44781\}, \{60, 1.23196\}, \{70, 1.06521\}, \{80, 0.924554\}, \{90, 0.791575\}\}$

$Y = g[X] = 2.2193 - 0.0160987 X$

Plot the least squares line in XY-space.



$Points = \{\{0, 2.25269\}, \{10, 2.03567\}, \{20, 1.87403\}, \{30, 1.70475\}, \{40, 1.62038\}, \{50, 1.44781\}, \{60, 1.23196\}, \{70, 1.06521\}, \{80, 0.924554\}, \{90, 0.791575\}\}$

$Y = g[X] = 2.2193 - 0.0160987\, X$

So the coefficients A and B are located at nodes (2,1) and (1), respectively:

$\mathbf{A} = \mathbf{g[x]}_{[[2,1]]}$

$\mathbf{B} = \mathbf{g[x]}_{[[1]]}$

-0.0160987
2.2193

Use $c = e^{B}$ and a = A to get the coefficients of $y = f_1[x] = \frac{L}{1 + c\, e^{a\, x}}$ back in the original xy-plane.

$\mathbf{c} = \mathbf{e}^{\mathbf{B}}$

$\mathbf{a} = \mathbf{A}$

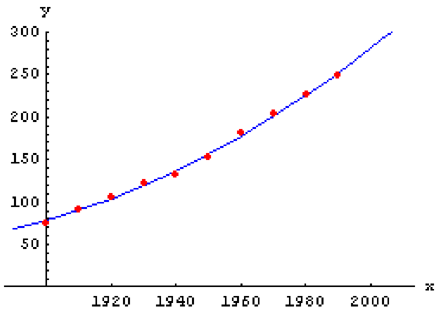
9.20092
-0.0160987

When we form the function, we must adjust "x" because we shifted the abscissas to the left. The actual form of the answer is a little different than what we original planned.

$Points = \{\{1900, 76.094\}, \{1910, 92.407\}, \{1920, 106.461\}, \{1930, 123.077\}, \{1940, 132.122\}, \{1950, 152.271\}, \{1960, 180.671\}, \{1970, 205.052\}, \{1980, 227.225\}, \{1990, 249.464\}\}$

$$y = f_1[x] = \frac{800}{1 + 9.20092\, e^{-0.0160987\,(-1900 + x)}}$$

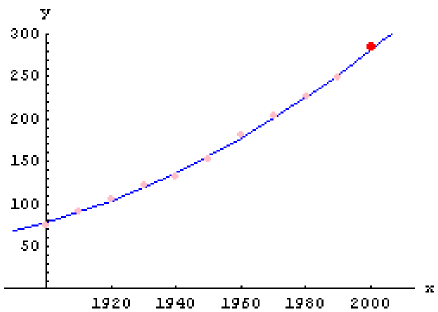
Now graph the function $y = f_1[x]$.



$\{(x_k, y_k)\} = \{\{1900, 76.094\}, \{1910, 92.407\}, \{1920, 106.461\}, \{1930, 123.077\},$
$\{1940, 132.122\}, \{1950, 152.271\}, \{1960, 180.671\}, \{1970, 205.052\}, \{1980, 227.225\}, \{1990, 249.464\}\}$

$$y = f_1[x] = \frac{800}{1 + 9.20092\, e^{-0.0160987\,(-1900+x)}}$$

So, estimate the population in 2000.



$$y = f_1[x] = \frac{800}{1 + 9.20092\, e^{-0.0160987\,(-1900+x)}}$$

$$y = f_1[2000] = 281.751$$