International Journal of Science Academic Research
Vol. 02, Issue 09, pp.3043-3049, September, 2021
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

INVESTIGATING THE BEST FITTING EMPIRICAL FUNCTION BETWEEN POWER AND EXPONENTIAL ONES

[1,*]Njipouakouyou Samuel and [2]Mahamat Saleh Daoussa Haggar

[1]Frmr senior lecturer, Faculty of Sciences, University of Dschang – CAMEROON
[2]Maître de Conférences, Dept de Mathématiques, Fac des Sciences Exactes et Appliquées, Laboratoire de Modélisation, Mathématiques, Informatique, Applications et Simulation (L2MIAS), Université de Ndjamena – TCHAD

Received 17th July 2021; **Accepted** 11th August 2021; **Published online** 30th September 2021

Abstract

This paper has shown the limitation of the least squares method for finding power and exponential functions from empirical data. It has shown the method of equal sums the best to solve this problem. Criteria of chosen a best fitting empirical function between two or more other ones has been elaborated in this paper. Modified formula of the least square method to minimize errors generated when transforming power and exponential functions to linear ones has been established.

Keywords: Experimental data, cloud of points, empirical function; least squares method, method of equal areas, method of equal sums, power function, exponential function, best fitting function, sum of the squares of the differences between empirical function at a nod and its corresponding experimental value.

1. INTRODUCTION

Up today, no instruments can directly give the functional analytic formula expressing a relationship between two or more variables. In this study, we shall limit ourselves to the case of two variables, the most encountered one. The searched function y = f(x) is unknown. Experimentally, some of its values, $y_i = f(x_i)$, can be found at certain nods x_i, i = 1, 2, 3, …, n. leading to a next table of experimental data, Table 1.

Table 1. Table of experimental points M_i (x_i, y_i).

x_i	x_1	x_2	…	x_n
yi	y_1	y_2	…	y_n

Using Table 1 could enable us finding an empirical function, y_{emp} = f(x), to replace as closer as possible the unknown y = f(x) under the recommendations that y_{emp} should be sample and keep almost all the physical aspect of the investigated phenomenon.
For finding y_{emp} the next procedure seems to be unavoidable.

At the first step, all the experimental points $M_i(x_i, y_i)$ should be placed in a coordinates plane xoy to obtain a cloud of points.

At the second step, the configuration of this cloud of points should be carefully examined to find out the form of the relationship between the variables, if linear, quadratic, power or exponential, between others. This seems to be the more crucial step as the physical aspects of the problem should be taken into consideration. Practical works confirm that apart the linear and quadratic forms, it is not always easy to take a decision between the power and exponential functions. One of the goals of this study is to make taking this decision easy.

The third step consists of using existing methods to find the analytical expression of y_{emp}. Between these methods, the most encountered ones are the least squares and equal areas methods. Unfortunately, these methods are not always easily applicable to all functions, as power and exponential ones, between others. The next goal of this paper is to show how to overcome this problem. So, modified least squares method for searching y_{emp} in the case of power and exponential functions and the criteria of chosen the best fitting one are presented in this paper. This work has five sections. The first and present one introduces the problem to be solved. The second one recalls the two most encountered methods, namely the least squares and equal areas methods. In the third section is presented the method of equal sums to be used to overcome some difficulties in the case of power and exponential functions given in their general forms. Section four is reserved to some concrete studies. The references are in section five.

2. Some recalls on the least squares and equal areas methods

Assume that the unknown function is analytically represented by a polynomial of the form:

$$y = a + bx + cx^2 + dx^3 + \dots \qquad (1)$$

Corresponding Author:** ***Njipouakouyou Samuel, FRMR senior lecturer, Faculty of Sciences, University of Dschang – CAMEROON

where a, b, c, d, … are coefficients to be found. If some concrete values are attributed to these coefficients, (1) is completely defined. So, (1) is a function of many variables, its coefficients, and we may write:

y = F(a, b; c, d, …). (2)

The principle of the least squares method recommends establishing the next function of the coefficients:

U(a, b, c, d, …) = $\sum_{i=1}^{n}(y - y_i)^2 = \sum_{i=1}^{n}(a + bx + cx^2 + dx^3 + \cdots - y_i)^2$. (3)

To find the coefficients of (1); we determine at each experimental point $M_i(x_i,y_i)$ the partial derivatives of (3) and we minimize them to zero. We obtain a system of equations to be solved for the coefficients. This system is:

$\frac{\partial U}{\partial a} = \sum_{i=1}^{n} 2(y - y_i)\frac{\partial(y-y_i)}{\partial a} = \sum_{i=1}^{n} 2(a + bx + cx^2 + dx^3 + \cdots - y_i) = 0,$ (4)
$\frac{\partial U}{\partial a} = \sum_{i=1}^{n} 2(y - y_i)\frac{\partial(y-y_i)}{\partial b} = \sum_{i=1}^{n} 2x(a + bx + cx^2 + dx^3 + \cdots - y_i) = 0,$ (5)
$\frac{\partial U}{\partial a} = \sum_{i=1}^{n} 2(y - y_i)\frac{\partial(y-y_i)}{\partial c} = \sum_{i=1}^{n} 2x^2\,(a + bx + cx^2 + dx^3 + \cdots - y_i) = 0,$ (6)
$\frac{\partial U}{\partial a} = \sum_{i=1}^{n} 2(y - y_i)\frac{\partial(y-y_i)}{\partial a} = \sum_{i=1}^{n} 2x^3\,(a + bx + cx^2 + dx^3 + \cdots - y_i) = 0,$ (7)
……………………………………………………………………………….. (8)

In equations (4) - (8), $x = x_i$
Solving system of equations (4)-(8) for the coefficients leads to their concrete values and to the searched y_{emp}.

The principle of the method of equal areas is based on the geometrical meaning of a defined integral
$\int_{x_1}^{x_n} y dx$ (9)

which is the area of the domain limited by the ox axis, the lines x_1 and x_n and the curve of y. Recall that we are searching (9) for a function y_{emp} whose curve is as closer as possible to the one of the unknown y. Consequently, the difference $y - y_{th} \approx 0$ and we may write:

$\int_{x_1}^{x_n}(y - y_i)\, dx = 0$ (10)

whence we deduce

$\int_{x_1}^{x_n} y dx = \int_{x_1}^{x_n} y_i dx$ (11)

which is the principle of the method of equal areas.

For this method, the domain of integration x_1-x_n should be divided into sub domains whose number should be equal to the number of coefficients to be found, then applying (11) to each sub domain leads to the system of equations to be solved for the coefficients.

Consider that the cloud of experimental points has led to either a power or an exponential function whose general equations are respectively:

$y_{emp} = a + bx^c$, (12)
$y_{emp} = a + bc^x$. (13)

Both functions have three coefficients to be determined. The least squares method leads to the next systems of equations:

- For the power function:
-

$\sum_{i=1}^{n}(a + bx_i^c - y_i) = 0,$ (14)
$\sum_{i=1}^{n} x_i^c\,(a + bx_i^c - y_i) = 0,$ (15)
$\sum_{i=1}^{n} x_i^c ln x_i\,(a + bx_i^c - y_i) = 0.$ (16)

- For the exponential function:

$\sum_{i=1}^{n}(a + bc^{x_i} - y_i) = 0,$ (17)
$\sum_{i=1}^{n} c^{x_i}\,(a + bc^{x_i} - y_i) = 0,$ (18)
$\sum_{i=1}^{n} c^{x_i} lnc\,(a + bc^{x_i} - y_i) = 0.$ (19)

It is clear that solving the systems of equations (14)-(16) and (17)-(19) is not evident and seems to be practically impossible. The conclusion remains the same with the method of equal areas.

One could try to transform (12) and (13) respectively to the next forms:

$\ln(y - a) = c\ln x,$ (20)
$\ln(y - a) = \ln b + x\ln c.$ (21)

It is clear that as the value of a is unknown, the systems of equations generated by (20)-(21) to be solved for a, b and c will be practically impossible. Thus, another method should be elaborated to overcome this difficulty.

3. The method of equal sums

The principles of the methods of equal sums and equal areas are similar. The main difference between them is that instead of integration, sum of y_{emp} at some nods x_i is assumed to be equal to the sum of their corresponding experimental ordinates y_i. Their common point is the division of the domain x_1-x_n into three sub domains, the number of coefficients to be found. Thus, during an experiment, it is recommended to take n multiple of three. So, we have $\frac{n}{3}$ =N. According to this method, the equations (12) and (13) will lead respectively to the next systems:

- For the power function:

$$Na + b\sum_{i=1}^{\frac{n}{3}} x_i^c = \sum_{i=1}^{\frac{n}{3}} y_i, \quad (22)$$
$$Na + b\sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} x_i^c = \sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} y_i, \quad (23)$$
$$Na + b\sum_{i=\frac{2n}{3}+1}^{n} x_i^c = \sum_{i=\frac{2n}{3}+1}^{n} y_i, \quad (24)$$

- For the exponential function:

$$Na + b\sum_{i=1}^{\frac{n}{3}} c^{x_i} = \sum_{i=1}^{\frac{n}{3}} y_i, \quad (25)$$
$$Na + b\sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} c^{x_i} = \sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} y_i, \quad (26)$$
$$Na + b\sum_{i=\frac{2n}{3}+1}^{n} c^{x_i} = \sum_{i=\frac{2n}{3}+1}^{n} y_i. \quad (27)$$

For system (22)-(24), coefficients a, b and c is progressively eliminated subtracting (22) from (23) and (23) from (24) to obtain a system of two equations in b and c; then diving member by member both equations leads to an equation of one variable c from which its value is found. Then other coefficients are found reverse way. These operations lead to the next equations:

$$b\left[\sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} x_i^c - \sum_{i=1}^{\frac{n}{3}} x_i^c\right] = \sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} y_i - \sum_{i=1}^{\frac{n}{3}} y_i, \quad (28)$$
$$b\left[\sum_{i=\frac{2n}{3}+1}^{n} x_i^c - \sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} x_i^c\right] = \sum_{i=\frac{2n}{3}+1}^{n} y_i - \sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} y_i; \quad (29)$$

and

$$\Phi(c) = \frac{\left[\sum_{i=\frac{2n}{3}+1}^{n} x_i^c - \sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} x_i^c\right]}{\left[\sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} x_i^c - \sum_{i=1}^{\frac{n}{3}} x_i^c\right]} = \frac{\sum_{i=\frac{2n}{3}+1}^{n} y_i - \sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} y_i}{\sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} y_i - \sum_{i=1}^{\frac{n}{3}} y_i}. \quad (30)$$

In the case of an exponential function the transformed equations are respectively:

$$b(c^{\alpha} - 1)\sum_{i=1}^{\frac{n}{3}} c^{x_i}, \quad (31)$$
$$b(c^{\alpha+\beta} - c^{\alpha})\sum_{i=1}^{\frac{n}{3}} c^{x_i}, \quad (32)$$

Dividing (32) by (31) leads to an equation of one variable easy to solve:

$$\Phi(c) = c^{\beta} = \chi. \quad (33)$$

In the case of power and exponential functions in their canonic form, i.e. without free term a, their equations are respectively;

$y = bx^c,$ (34)
$y = bc^x.$ (35)

Both equations have two coefficients b and c to be found solving a system of two equations. Transforming both members of these equations using logarithm leads respectively to:

$\ln y = \ln b + c\ln x,$ (36)
$\ln y = \ln b + x\ln\alpha.$ (37)

Putting $\ln y = Y$, $\ln x = X$, $\ln b = B$, $\ln c = C$, (36) and (37) become:

$Y = B + cX,$ (38)
$Y = B + Cx,$ (39)

which are two linear equations to be solved for the coefficients first in the coordinates system XoY, and then the results transferred to the coordinates system xoy.

In general, introducing the logarithm generates some round off error Δ ignored by many researchers during their investigations. This could seriously affect the results. One should take it into account. Letting Δ the error between ordinates y_{emp} and unknown y leads to:

$Y_{emp} - y = \Delta,\ y_{emp} = y + \Delta.$ (40)

When passing to the new ordinate $Y = f(y)$, we have:

$Y_{emp} = f(y_{emp}) = f(y + \Delta) \approx f(y) + f'(y)\,\Delta = Y + f'(y)\,\Delta,$ (41)

Whence

$\Delta = \frac{Y_{emp} - Y}{f'(y)}$ (42)

and the sum of the squares of the differences between y_{emp} and experimental y_i while taking into consideration the error generated when transforming (34) and (35) to the linear forms is:

$\sum_{i=1}^{n} \frac{1}{f'(y)} (Y_{emp,i} - Y_i)^2$. (43)

As $f(y) = \ln y$, we have $f'(y) = \frac{1}{y}$, and $\frac{1}{f'(y)} = y$.. Therefore, formula (43) becomes:

$\sum_{i=1}^{n} y_i\,(B + cX_i - Y_i)^2$. (44)

And the system of equations to be solved for B and c is obtained minimizing to zero the next expressions:

$\sum_{i=1}^{n} y_i\,(B + cX_i - Y_i)$. (45)
$\sum_{i=1}^{n} X_i\, y_i\,(B + cX_i - Y_i)$. (46)

which are developed to the following ones:

$B\sum_{i=1}^{n} y_i + c\sum_{i=1}^{n} X_i\, y_i = \sum_{i=1}^{n} Y_i\, y_i,$ (47)
$B\sum_{i=1}^{n} X_i\, y_i + c\sum_{i=1}^{n} X_i^2\, y_i = \sum_{i=1}^{n} X_i\, Y_i\, y_i.$ (48)

It may happen that we have to choose the best fitting y_{emp} between two empirical functions issued from the treatment of a same experimental data. To solve this problem we calculate for each y_{emp} the sum of the squares of the differences between $y_{emp,i}$ and y_i:

$\sum_{i=1}^{n} (y_{emp,i} - y_i)^2$. (49)

The empirical function for which (49) is minimal is the best fitting one.

4. Some concrete cases

The results of an experiment are presented in Table 2. Find y_{emp} to be as closer as possible to the unknown function $y = f(x)$.

Table 2. Experimental data

x_i	0	2	4	6	8	10	12	14	16	18	20	22
y_i	6	7	5	5	6	7	9	13	17	22	31	37

Plotting the points $M_i(x_i,y_i)$ in the plane xoy and analyzing its configuration lead to either a power or exponential functional relationship, equations (12) and (13). We have 12 points, a multiple of 3. We form three sub domains of 4 points each.
Case of the power function:

The equations of the system are:

First sub domain: $4a + b(0^c + 2^c + 4^c + 6^c) = 6 + 7 + 5 + 5 = 23$, (50)

Second sub domain: $4a + b(8^c + 10^c + 12^c + 14^c) = 6 + 7 + 9 + 13 = 35$, (51)

Third sub domain: $4a + b(16^c + 18^c + 20^c + 22^c) = 17 + 22 + 31 + 37 = 107$, (52)

(50)-(52) is the system of three equations to be solved for a, b and c using the method of equal sums. Eliminate a from these equations by the operations (51)-(50) and (52)-(51) leading to the next system of two equations in b and c:

$b[(8^c + 10^c + 12^c + 14^c) - (0^c + 2^c + 4^c + 6^c)] = 12$, (53)

$b[(16^c + 18^c + 20^c + 22^c) - (8^c + 10^c + 12^c + 14^c)] = 72$. (54)

Next, eliminate b dividing (54) by (53):

$\frac{[(16c + 18c + 20c + 22c) - (8c + 10c + 12c + 14c)]}{[(8c + 10c + 12c + 14c) - (0c + 2c + 4c + 6c)]} = \frac{72}{12} = 6 = \Phi(c)$. (55)

The problem is to find c for which $\Phi(c) = 6$. Giving c different values, we have:

$\Phi(1) = \frac{76-44}{44-12} = 1$, $\Phi(2) = \frac{1464-504}{504-56} = 2.14$,

$\Phi(3) = \frac{28576-5984}{5984-288} = 3.97$, $\Phi(4) = \frac{564768-73248}{73248-1568} = 6.86$,

Plotting the curve of $\Phi(c)$ and interpolating lead to the approximation $c \approx 3.75$. Replacing c by its value in (53) leads to $b \approx 0.0003153$. Replacing b and c by their values in (50) leads to $a \approx 5.67$. Whence the power function:

$y_{emp} = 5.67 + 0.0003153x^{3.75}$. (56)

Case of the exponential function:

In the first, second and third sub domains, the equations of the system are respectively:

$4a + b(c^0 + c^2 + c^4 + c^6) = 23$, (57)

$4a + b(c^8 + c^{10} + c^{12} + c^{14}) = 35$, (58)

$4a + b(c^{16} + c^{18} + c^{20} + c^{22}) = 107$, (59)

Next are the subtractions to eliminate a:

(58)-(57) $= b(c^8 - 1)(c^0 + c^2 + c^4 + c^6) = 12$, (60)
(59)-(58) $= b(c^{16} - c^8)(c^0 + c^2 + c^4 + c^6) = 72$, (61)

Dividing (61) by (60) to eliminate b leads to

$\frac{(61)}{(60)} = \Phi(c).= c^8 = 6$, (62)

whence $c \approx 1.251$. And proceeding by analogy to the above case, we have $b \approx 0.2712$ and $a \approx 5.15$. Therefore, the searched empirical exponential function is:

$y_{emp} = 5.15 + 0.2712(1.251)^x$. (63)

Between (56) and (63), only one should better fit the phenomenon. Finding it leads to the computations of the sum $\sum_{i=1}^{12}(y_{emp,i} - y_i)^2$ for each formula, confer to Table 3.

Table 3. Intermediary computations. In this table (56) and (63) just mean formulas (56) and (63) were used in computations. 56 and 63 are referred to the corresponding formulas

x_i	y_i	(56)	$(y_{emp,i} - 56)^2$	(63)	$(y_{emp,i} - 63)^2$
0	6	5.67	0.109	5.421	0.335
2	7	5.674	1.758	5.574	2.033
4	5	5.727	0.529	5.814	0.663
6	5	5.931	0.867	6.190	1.416
8	6	6.438	0.192	6.777	0.604
10	7	7.443	0.196	7.696	0.484
12	9	9.183	0.033	9.135	0.018
14	13	11.932	1.141	11.386	2.605
16	17	16.002	0.996	14.909	4.372
18	22	21.739	0.068	20.423	2.487
20	31	29.525	2.176	29.052	3.795
22	37	39.774	7.695	42.557	30.880
$\sum$			15.760		49.692

Table 3 shows that$(y_{emp,i} - 56)^2$ is less than$(y_{emp,i} - 63)^2$, consequently, (56) is the best fitting formula compare to (63). So we should consider (56) for further work. The second case concerns the power function in its canonic form. Consider Table 4 of experimental data to find y_{emp} which should better describe the investigated phenomenon.

Table 4. Experimental data

x_i	1	2	3	4	5	6	7	8	9	10
y_i	0.5	0.5	1	2	4	5	6	8	12	15

Analyzing the configuration of experimental points $M_i(x_i,y_i)$ plotted in xoy plane leads to a power function, therefore, its equation is $y = bx^c$ to be transformed to a linear form using logarithm:

$\ln y = \ln b + c\ln x$ or $Y = B + cX$, (64)
$Y = \ln y$, $X = \ln x$, $B = \ln b$.

Applying the least squares method to (64) leads to the next system of two equations to be solved for b and c:

$nB + c\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i,$ (65)
$B\sum_{i=1}^{n} X_i + c\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i$ (66)

Replacing these sums by their values leads to the concrete system:

$10B + 6.559c = 4.936,$
$6.559B + 5.214c = 4.732,$

whence B = -0.581 and b = 0.262; c = 1.639
and the searched empirical function is:

$y_{emp} = 0.262\, x^{1.939}.$ (67)

Taking into account the error generated when introducing the logarithm, we have:

$B\sum_{i=1}^{n} y_i + c\sum_{i=1}^{n} X_i y_i = \sum_{i=1}^{n} Y_i y_i,$ (68)
$B\sum_{i=1}^{n} X_i y_i + c\sum_{i=1}^{n} X_i^2 y_i = \sum_{i=1}^{n} X_i Y_i y_i,$ (69)

To go faster in finding the above sums, the next table of intermediary computations is proposed, Table 5.

Table 5 Intermediary computations

i	x_i	y_i	X_i	Y_i	X_iy_i	Y_iy_i	$Y_iX_i^2$	$y_iX_iY_i$
1	1	0.5						
2	2	0.5						
3	3	1						
4	4	2						
5	5	4						
6	6	5						
7	7	6						
8	8	8						
9	9	12						
10	10	15						
$\sum$		54	-	4.936	47.259	48.685	42.706	45.178

Replacing these sums in (68)-(69) leads to the concrete system of equations

$54B + 47.259c = 48.685,$
$47.259B + 42.706c = 45.178,$

whence B = -0.7647 and b = 0.172; c ≈ 1.9.

Consequently, the searched empirical function is

$y_{emp} = 0.172x^{1.9}.$ (70)

We verify that formula (70) gives better results than (67). For this purpose, we build Table 6.

Table 6. For verification of the fitness of (70) compared to (67)

x_i	y_i	(67)	$(y_i – 67)^2$	(70)	$(y_i – 70)^2$
1	0.5	0.262	0.057	0.172	0.108
2	0.5	0/816	0.100	0.642	0.020
3	1	1.586	0.343	1.387	0.150
4	2	2.541	0.293	2.396	0.157
5	4	3.664	0.113	3.661	0.115
6	5	4.940	0.004	5.176	0.031
7	6	6.359	0.129	6.938	0.880
8	8	7.915	0.007	8.941	0.885
9	12	9.601	5.755	11.184	0.666
10	15	11.410	12.888	13.662	1.790
$\sum$			19.689		4.802

The sum generated by (70) is almost 5 times less the one generated by (67) Whence the conclusion that the introduced modification has considerably improved the results and must be recommended for operational works.

5. Conclusion

No doubt that the least squares method plays a very important role in the treatment of experimental data. This method goes well with linear and quadratic functions. Concerning other types of functions, one should be careful, particularly with the power and exponential functions when given in their general forms, i.e. with free term. For such functions the method of equal sums simplifies the computations and therefore, is very recommended in operational works. This paper also attracts attention of researchers on the errors generated when transforming canonic power and exponential functions to linear forms as it could seriously affect the results. Criteria of chosen the best fitting empirical function between two or more other ones has been established.

Reference

Melentev P. V. Approximate computations. Gosudarstvennoe Izdatelsvo Phiziko-Mathematicheskoe Literaturi. Moscow 1962. Pp 320-366.

****