Southern Federal University, Russia, 3 June 2020

Mathematical Methods in Economics

Least Squares Approximation Method

Prof. Dr. Vesselin Gushev

Faculty of Mathematics and Informatics, University of Sofia, Bulgaria

Sometimes a priori we know that the function f we are considering has the following properties:

- The function f depends on $n+1$ parameters $a_0, \ldots, a_n$ $\left(\text{ex. } \sum_{k=0}^{n} a_k x^k, \quad \prod_{k=0}^{n} \sin a_k x, \quad \sum_{k=0}^{n} e^{a_k x} \text{ etc.}\right)$;

- We can calculate the values of the function f with a given precision in a set of nodes (but in practice actually calculating the value of the function in some point may be a costly or time consuming procedure).

Our goal is to calculate the parameters $a_0, \ldots, a_n$ as precisely as possible, based on the data:

$$f(x_1), f(x_2), \ldots, f(x_m); \quad m > n + 1 \ .$$

The values, listed in this information are almost never the exact values of the function f, but are their close approximations.

Let $y = f(x)$ be linear

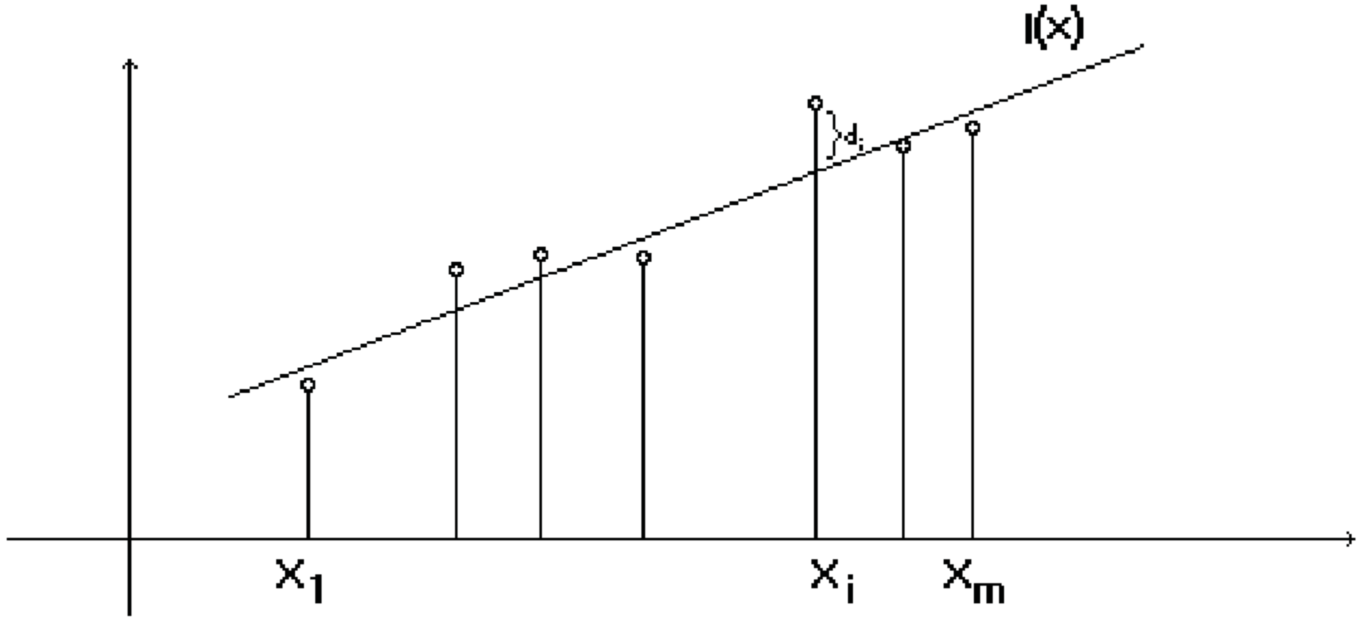
$$f(x) = Ax + B \ ,$$

for some A and B.

The approximations of the values of f

$$f_i = f(x_i), \ i = 1, \dots, m,$$

are shown on the figure below.



Because of flaws in the measuring of the values or the experiment itself, the points (x_i, f_i), $i = 1, \ldots, m$, do not lie on a line.

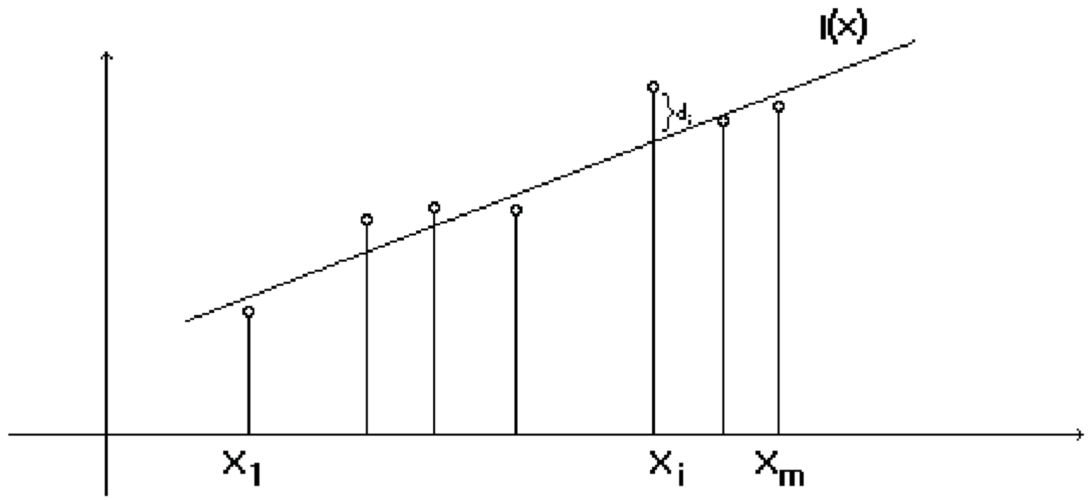
As $f(x)$ is linear, then the question arises:

Which line do we choose to represent the data?

There are numerous lines to choose from. One way would be to fix two points $(x_i, f_i), (x_j, f_j)$ and to take the line l through them as a representation of f. The problem is that such a choice would be arbitrary.

Let's do something a bit more methodical.

So, we are looking for a function in the form $l(x) = Ax + B$.



We denote by d_i the deviation of f_i from the value $l(x_i)$, predicted by l in the node x_i

$$d_i := f_i - (Ax_i + B) , \qquad i = 1, \ldots, m .$$

There are a few reasonable approaches to choosing the parameters A and B of l.

1) The parameters A and B are chosen to minimize the value

$$\max_{1 \leqq i \leqq m} |d_i| \ .$$

So we try to minimize the maximal distance between f and l in the nodes $x_1, \ldots, x_m$.

This is a viable strategy, but it leads to a non-linear problem (the minimized function is non-linear with respect to A and B).

2) The parameters A and B are chosen to minimize the value

$$\sum_{i=1}^{m} |d_i| \ .$$

The same difficulties arise as in the previous case. Solving such non-linear problems was almost impossible before the invention of modern computers.

3) The parameters A and B are chosen to minimize the value

$$S(A,B) := \sum_{i=1}^{m} d_i^2 \ .$$

This method for finding unknown parameters of a function, based on its data is called The Least Squares Method.

Necessary conditions (which are sufficient too) for a minimum of

$$S(A,B) = \sum_{i=1}^{m} [f_i - (Ax_i + B)]^2$$

are

$$\frac{\partial S}{\partial A} = 0 \;\Rightarrow\; \sum_{i=1}^{m} [f_i - (Ax + B)]x_i = 0,$$

$$\frac{\partial S}{\partial B} = 0 \;\Rightarrow\; \sum_{i=1}^{m} [f_i - (Ax + B)] = 0.$$

So that strategy finally leads to a linear system for A and B.

Now we will consider The Least Squares Method in a more general form:

Denote $\Omega_n = \left\{ F(x, a_0, \ldots, a_n) = \sum_{i=0}^{n} a_i \varphi_i(x) : a_i \in I_i, i = 0, \ldots, n \right\}$, where $\varphi_0(x), \varphi_1(x), \ldots, \varphi_n(x)$, $i = 0, \ldots, n$ are linearly independent functions.

The data $f_1, \ldots, f_m$ cold be considered as values of a function $f(x) \in l_2(x_1, x_2, \ldots, x_m)$, respectively in $x_1, x_2, \ldots, x_m,$, where $m > n + 1$.

Definition. We will say that $F(x, a_0, \ldots, a_n)$ is an approximation of the data $f_1, \ldots, f_m$ using the least squares method, if the parameters $a_0, \ldots, a_n$ are chosen to minimize the sum

$$\sum_{i=1}^{m} \mu_i [F(x_i, a_0, \ldots, a_n) - f_i]^2 ,$$

where $\{\mu_i\}_i^m$ are fixed positive numbers, called "weights".

Our goal is to find an approximation of the form

$$F(x, a_0, \ldots, a_n) = \sum_{i=0}^{n} a_i \varphi_i(x)$$

of f using the least squares method, based on the data

$$f_i = f(x_i), \quad i = 1, \ldots, m$$

.

Let $\{\mu_i\}, \quad i = 1, \ldots, m$ be the set of weights.

Then we have to minimize

$$\Phi(a_0, \ldots, a_n) := \sum_{i=1}^{m} \mu_i \left[f_i - \sum_{k=0}^{n} a_k \varphi_k(x_i) \right]^2 \quad .$$

with respect to the parameters $a_0, a_1, \ldots, a_n$.

In fact $\Phi^2(a_0, \ldots, a_n)$ is the distance between f and $F(x, a_0, \ldots, a_n)$ in the Hilbert space $l_2(x_1, x_2, \ldots, x_m)$ of functions, defined in nodes $x_1, \ldots, x_m$, with the following inner product

$$[h, g] := \sum_{i=1}^{m} \mu_i h(x_i) g(x_i) \ .$$

This inner product induces the norm

$$\|g\| := \left\{ \sum_{i=1}^{m} \mu_i g^2(x_i) \right\}^{1/2} ,$$

which in turn induces the distance

$$\rho(h,g) = \left\{ \sum_{i=1}^{m} \mu_i [h(x_i) - g(x_i)]^2 \right\}^{\frac{1}{2}} .$$

So our function $\Phi(a_0, \ldots, a_n)$ represents the distance between f and $F(x, a_0, \ldots, a_n)$.

Then the least squares method leads to a problem of best approximation in the Hilbert space $l_2(x_1, x_2, \ldots, x_m)$.

Now from the theory follows that the parameters $a_0, \ldots, a_n$ are the solutions of the linear system (*):

$$a_0[\varphi_0, \varphi_0] + a_1[\varphi_1, \varphi_0] + \ldots + a_n[\varphi_n, \varphi_0] = [f, \varphi_0]$$

$$(*) \qquad a_0[\varphi_0, \varphi_1] + a_1[\varphi_1, \varphi_1] + \ldots + a_n[\varphi_n, \varphi_1] = [f, \varphi_1]$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_0[\varphi_0, \varphi_n] + a_1[\varphi_1, \varphi_n] + \ldots + a_n[\varphi_n, \varphi_n] = [f, \varphi_n] \ .$$

That is a linear system of $n+1$ equations with $n+1$ variables. Let $D(\varphi_0, \ldots, \varphi_n)$ be the determinant of (*).

We have

$$D(\varphi_0, \ldots, \varphi_n) := \det \begin{bmatrix} [\varphi_0, \varphi_0] & [\varphi_1, \varphi_0] & \cdots & [\varphi_n, \varphi_0] \\ [\varphi_0, \varphi_1] & [\varphi_1, \varphi_1] & \cdots & [\varphi_n, \varphi_1] \\ \vdots & \vdots & \ddots & \vdots \\ [\varphi_0, \varphi_n] & [\varphi_1, \varphi_n] & \cdots & [\varphi_n, \varphi_n] \end{bmatrix} .$$

This is Gram determinant, which is not zero, because the functions $\varphi_0, \ldots, \varphi_n$ are linearly independent.

Thus the system (*) has a unique solution $a_0, \ldots, a_n$.

Now we will consider a problem of approximating a function with algebraic polynomials of degree n in the set of nodes

$$x_1 < \ldots < x_m \quad (m > n + 1).$$

In this case we have $\varphi_k(x) = x^k$ and the polynomial is

$$p(x) = a_0 + a_1 x + \ldots + a_n x^n.$$

Then the linear system of $a_0, \ldots, a_n$ has the form:

$$a_0[\varphi_0(x), \varphi_k(x)] + a_1[\varphi_1(x), \varphi_k(x)] + \cdots + a_n[\varphi_n(x), \varphi_k(x)] = [f(x), \varphi_k(x)]$$

$$(k = 0, \ldots, n).$$

or

$$a_0 \sum_{i=1}^{m} x_i^k + a_1 \sum_{i=1}^{m} x_i^{k+1} + \cdots + a_n \sum_{i=1}^{m} x_i^{k+n} = \sum_{i=1}^{m} f(x_i)\, x_i^k$$

$$(k = 0, \ldots, n).$$

The process of solving a linear system is not a quick one, but it can be evaded by choosing a proper basis in the space of algebraic polynomials of degree n.

If we present p in the form

$$p(x) = b_0 P_0(x) + \cdots + b_n P_n(x),$$

where $\{P_k(x)\}$ form an orthogonal system of polynomials in $l_2(x_1, \ldots x_m)$ with weights $\{\mu_i\}$, then the system will be reduced to a diagonal one:

$$b_k [P_k(x), P_k(x)] = [P_k(x), f(x)]$$

$$(k = 0, \ldots, n).$$

Now the parameters b_k can be determined directly.

Southern Federal University, Russia, 3 June 2020

Thank you for your attention!