LEAST SQUARES

OLIVIER CASTÉRA

Abstract. The method of least squares is based on a solid probabilistic foundation.

Contents

1 Justification of the least squares method

Theorem. *When measurement errors follow a normal distribution, for a set of observed values $(y_1, y_2, \ldots, y_n)$ of a function $y = \varphi(x)$ to be determined to be the most probable, the function must be chosen such that the sum of the squares of the deviations of the observed values from $\varphi(x)$ is minimal:*

$$d\sum_{i=1}^{n}[y_i - \varphi(x_i)]^2 = 0$$

Proof. Let $y = \varphi(x)$ be the exact expression of the true relationship between y and x. The experimental points deviate from this relationship due to unavoidable measurement errors. The central limit theorem[1] shows that measurement errors generally follow a normal distribution[2]. Assume this is the case, and choose a value x_i of the argument for which we perform a measurement. The result of this measurement is the event $Y_i = y_i$, where Y_i is a random variable distributed according to a normal law with mathematical expectation $\varphi(x_i)$ and standard deviation σ_i. The expectation $\varphi(x_i)$ is the result that would be obtained if there were no measurement error, and σ_i characterizes the measurement error. Assume the measurement error is constant at every point:

$$\forall i \ \sigma_i = \sigma$$

The probability density of the variable Y_i is:

$$f_i(y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-[y_i - \varphi(x_i)]^2/(2\sigma^2)}$$

Assume that the experiment, which in our example is a series of measurements, has realized the event consisting of the random variables $(Y_1, Y_2, \ldots, Y_n)$ taking the values $(y_1, y_2, \ldots, y_n)$.

Date: December 14, 2025.

[1]See Central Limit Theorem.pdf

[2]See Normal Distribution.pdf

Since the variables Y_i are continuous, the probability of each event $Y_i = y_i$ is zero, which is why we consider the corresponding probability elements:

$$f_i(y_i)dy_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-[y_i-\varphi(x_i)]^2/(2\sigma^2)} dy_i$$

Since the measurements are independent, the probability that the system of random variables $(Y_1, Y_2, \dots, Y_n)$ takes the set of values each lying in the interval $(y_i, y_i + dy_i)$ is equal to the product of the probability elements:

$$\begin{aligned}\prod_{i=1}^{n} f_i(y_i)dy_i &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-[y_i-\varphi(x_i)]^2/(2\sigma^2)} dy_i \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{[-1/(2\sigma^2)]\sum_{i=1}^{n}[y_i-\varphi(x_i)]^2} \prod_{i=1}^{n} dy_i\end{aligned}$$

For this probability to be maximal, it is necessary that:

$$\sum_{i=1}^{n}[y_i - \varphi(x_i)]^2$$

be minimal. □

2 Numerical characteristics of systems of two variables

Definition 2.1. *The initial moment of order k, s of the system of random variables (X, Y) is defined as the mathematical expectation of the product X^kY^s:*

$$\alpha_{k,s}[X^kY^s] \triangleq M[X^kY^s]$$

For discrete random variables:

$$\alpha_{k,s}[X^kY^s] \triangleq \sum_{i=1}^{n}\sum_{j=1}^{m} x_i^k y_j^s p_{ij}$$

For continuous random variables:

$$\alpha_{k,s}[X^kY^s] \triangleq \iint_{-\infty}^{+\infty} x^k y^s f(x, y)dxdy$$

Definition 2.2. *The centered moment of order k, s of the system of random variables (X, Y) is defined as the mathematical expectation of the product $\mathring{X}^k\mathring{Y}^s$:*

$$\mu_{k,s}[X^kY^s] \triangleq M[\mathring{X}^k\mathring{Y}^s]$$

For discrete random variables:

$$\mu_{k,s}[X^kY^s] \triangleq \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i - m_x)^k (y_j - m_y)^s p_{ij}$$

For continuous random variables:

$$\mu_{k,s}[X^kY^s] \triangleq \iint_{-\infty}^{+\infty} (x - m_x)^k (y - m_y)^s f(x, y)dxdy$$

Definition 2.3. *The covariance of the random variables X, Y is the mixed centered moment of order two $\mu_{1,1}$:*

$$K_{xy} \triangleq M[\mathring{X}\mathring{Y}]$$

For discrete random variables:

$$K_{xy} \triangleq \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - m_x)(y_i - m_y)p_{ij}$$

For continuous random variables:

$$K_{xy} \triangleq \iint_{-\infty}^{+\infty}(x - m_x)(y - m_y)f(x, y)dxdy$$

3 Statistical characteristics of a distribution

Every numerical characteristic[3] of the random variable X has a corresponding statistical characteristic.

Definition 3.1. *The arithmetic mean or statistical mean of the discrete random variable X is the average of all observed values x_i of this variable. It is denoted $M^*[X]$ or m_x^*.*

$$M^*[X] \triangleq \frac{1}{n}\sum_{i=1}^{n} x_i$$

Definition 3.2. *The statistical variance of the discrete random variable X is the statistical mean of the squares of the deviations of the observed values from the statistical mean. It is denoted $V^*[X]$ or V_x^* or $D^*[X]$ or D_x^*.*

$$V^*[X] \triangleq \frac{1}{n}\sum_{i=1}^{n}(x_i - m_x^*)^2$$

Definition 3.3. *The initial statistical moment of order s of the discrete random variable X is the function $\alpha_s^*[X]$ defined by:*

$$\alpha_s^*[X] \triangleq \frac{1}{n}\sum_{i=1}^{n} x_i^s$$

Definition 3.4. *The centered statistical moment of order s of the discrete random variable X is the function $\mu_s^*[X]$ defined by:*

$$\mu_s^*[X] \triangleq \frac{1}{n}\sum_{i=1}^{n}(x_i - m_x^*)^s$$

[3] See Normal Distribution.pdf

Definition 3.5. *The statistical covariance of the discrete random variables X and Y is the function K^*_{xy} defined by:*

$$K^*_{xy} \triangleq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i - m^*_x)(y_i - m^*_y)$$

4 Application

An experiment has provided a set of values $(x_i, y_i)_{i=1,\ldots,n}$. We assume that the measurement errors follow a normal distribution with constant standard deviation, and that the observed relationship is linear. We seek the parameters a and b of the line

$$y = ax + b$$

that best represent the experimental relationship:

$$\begin{aligned} d\sum_{i=1}^{n} [y_i - \varphi(x_i)]^2 &= \sum_{i=1}^{n} d[y_i - \varphi(x_i)]^2 \\ &= \sum_{i=1}^{n} d[y_i^2 - 2y_i\,\varphi(x_i) + \varphi^2(x_i)] \\ &= \sum_{i=1}^{n} -2y_i\,d\varphi(x_i) + 2\varphi(x_i)\,d\varphi(x_i) \\ &= \sum_{i=1}^{n} 2[\varphi(x_i) - y_i]\,d\varphi(x_i) \\ &= \sum_{i=1}^{n} 2[\varphi(x_i) - y_i]\,[\partial_a\varphi(x_i)da + \partial_b\varphi(x_i)db] \\ &= 0 \end{aligned}$$

whence the system

$$\begin{cases} \displaystyle\sum_{i=1}^{n} 2[\varphi(x_i) - y_i]\,\partial_a\varphi(x_i)da = 0 \\ \displaystyle\sum_{i=1}^{n} 2[\varphi(x_i) - y_i]\,\partial_b\varphi(x_i)db = 0 \end{cases}$$

$$\begin{cases} \displaystyle\sum_{i=1}^{n} [(ax_i + b) - y_i]\,x_i = 0 \\ \displaystyle\sum_{i=1}^{n} [(ax_i + b) - y_i] = 0 \end{cases}$$

$$\begin{cases} \displaystyle a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i x_i = 0 \\ \displaystyle a\sum_{i=1}^{n} x_i + nb - \sum_{i=1}^{n} y_i = 0 \end{cases}$$

Dividing by n and using the definitions from the previous chapter,

$$\begin{cases} a\alpha_2^*[X] + bm_x^* - \alpha_{1,1}^*[X,Y] = 0 \\ am_x^* + b - m_y^* = 0 \end{cases} \Rightarrow \begin{cases} b = m_y^* - am_x^* \\ a\alpha_2^*[X] + (m_y^* - am_x^*)m_x^* - \alpha_{1,1}^*[X,Y] = 0 \end{cases}$$

$$\begin{cases} a\left\{\alpha_2^*[X] - (m_x^*)^2\right\} = \alpha_{1,1}^*[X,Y] - m_x^* m_y^* \\ b = m_y^* - am_x^* \end{cases} \Rightarrow \begin{cases} a = \dfrac{\alpha_{1,1}^*[X,Y] - m_x^* m_y^*}{\alpha_2^*[X] - (m_x^*)^2} \\ b = m_y^* - am_x^* \end{cases}$$

whence, using centered moments:

$$\begin{cases} a = \dfrac{K_{xy}^*}{V_x^*} \\ b = m_y^* - am_x^* \end{cases}$$

Email address : o.castera@free.fr
Website : https://sciences-physiques.neocities.org