DOI:
http://dx.doi.org/10.15448/1983-4012.2018.1.31461

PROBABILITIES IN EXPERIMENTAL PHYSICS: EPISTEMIC LESSONS AND CHALLENGES

Probabilidades em Física Experimental: Lições Epistêmicas e Desafios

Thales Borrely *

Diana Taschetto **

Abstract: There is one way with which Nature responds to the questions we direct her about the correctness of our understanding of her ways: by means of experiments. In this paper, the pivotal role probability theory plays in experimental physics is presented: it allows us to combine observations that are seemingly analytically incompatible. The main concepts used for such task are introduced and explained. A brief historical sketch of the development of some of such concepts is drawn and it was used as a case-study to defend the position that physics and philosophy are interlinked affairs. Some philosophical consequences of how the intrinsically probabilistic character of experimentation reverberates in our epistemic access of the world are also drawn.
Keywords: Probability; Measurement; Philosophy of Experimental Physics.

Resumo: É por meio de experimentos que a Natureza responde às questões que a ela direcionamos sobre a adequação de nosso entendimento sobre seu funcionamento. O papel central que a teoria de probabilidades desempenha em física experimental é, neste artigo, apresentada: esta nos permite combinar observações que parecem, em uma primeira leitura, incompatíveis. Os conceitos principais utilizados para fazê-lo são introduzidos e explicados. Uma breve discussão histórica do desenvolvimento de alguns dos mencionados conceitos é indicada e utilizada como estudo de caso para dar suporte à posição que defende que física e filosofia são atividades entrelaçadas. Algumas consequências filosóficas acerca das implicações do caráter intrinsecamente probabilístico da experimentação para o nosso acesso epistêmico do mundo são, também, apresentadas.
Palavras-chave: Probabilidade; Medição; Filosofia da Física Experimental.

* Department of Material Physics, University of São Paulo, IF-USP, New Semiconductor Materials Laboratory. Master of Science by and PhD student at the University of São Paulo. Funded by CNPq. E-mail: thales.santos@usp.br.

** Philosophy Department, University of São Paulo, FFLCH-USP; Department of Mathematical Physics, University of São Paulo, IF-USP. Master of Philosophy by and PhD student at the University of São Paulo. Funded by CAPES. E-mail: dtaschetto@usp.br.

1. Introduction

"What is probability?" is a question, one could argue, no one can answer. This statement is supported by the vast literature produced by philosophers and scientists while struggling (A) to come up with a non-circular definition and (B) to interpret the meaning of "probability" (Sklar, 1979). The existence to this day of such debates is, by itself, astonishing if we consider that probability theory had its bases built (at least) more than three hundred years ago and, since then, has been an indisputably successful tool for assessing the world. Such success makes of probability an object of interest for a wide range of academics of different areas. Philosophers, for instance, often write about its interpretation and definition (Earman, 1992). On the other hand, scientists use probability on a daily basis to interpret experiments, to solve problems and even to forge core concepts of theories (Von Weizsäcker, 1973; Sachkov, 1928). Looking at this scenario, it is clear that, from the debate, an opportunity arises: the discussion about the meaning and definition of probability is a battlefield where philosophers and scientists fight side by side. Consequently, the history of probability is an experiment where epistemologists can find high quality data; it is a rich source of information about the intersection between philosophy and science. From this perspective, this article aims to present (a) a summary of how scientists use probability for measuring physical quantities, including constant parameters of functions, and (b) a brief description of the history of the development of part of such mathematical tools. Once (a) and (b) are completed, we shall advance towards (c) identifying which features of the presented facts are of interest for academics working in the intersection between science and philosophy, and which questions these features may help answering.

2. Measurement and uncertainty

Our first task is to present the *status quo* regarding how scientists measure physical quantities and what role probability plays in measurements. The content of this topic is no more than a summary of what physics undergraduate students usually learn in their first couple of experimental physics courses. Same facts can be found in any introductory textbook of data analysis (see, for instance, Vuolo, 1996; Taylor, 1997).

To begin, suppose that, in the context of an experiment, a physical quantity, call it w, must be measured. We can imagine, for simplicity, that w is the size of an object. To do so, it is necessary to choose a measuring instrument: a measuring-tape, a rule, a caliper rule or a micrometer, for instance. After choosing the instrument, the measurer measures the object and gets a result w_1. However, as strange as this may seem for those not used to experimentation, if another measurement is performed on the same object, a different result w_2 will be obtained. Repeating the procedure N times will yield a set

($w_1, w_2, ..., w_n$) of measurements.[1] We can represent this set in the form of a histogram, *i.e.*, a column graph representing how many results were obtained between the real numbers[2] ($x_1, x_1+\Delta x$), ($x_1+\Delta x, x_1+2\Delta x$) and so on. Note that this distribution enables a quick calculation of relative frequencies. The relative frequency of a measurement being within a certain interval is the height of the column relative to this interval over the sum of the heights of all columns.

We can abstract the situation to a continuous case with infinitesimally small intervals and represent the histogram by a curve. In this case, the relative frequency of an interval will be yield by the area under the curve delimited by such interval over the total area under the curve, *i.e.*, the relative frequency equals the relative area. If we take relative frequencies to represent probabilities, then the area under the curve delimited by an interval is the probability of a measurement being within that interval. For this reason, this kind of curve is called *probability density function* (PDF). Now, we can finally understand that the final result of our experiment is not a numerical value from the set of real numbers, as one would expect. It is, rather, a *distribution* we commonly represent by a PDF.

There is, nonetheless, a way of representing (at least partially) the PDF with a couple of real numbers. If the curve is symmetric, for instance, it is intuitively noticeable that we can represent the distribution by its central value (mean) plus a number representing the distribution's "width". In fact, this is exactly what scientists do. Usually, experimental results are presented in the form: $\overline{w} \pm \sigma$. $\overline{w}$ is the arithmetic mean of w over the N measurements. σ, called "standard deviation", is nothing more than an indicative of how far the data fluctuate around the mean. One way of grasping the meaning of σ is the following: it is expected that, if another measurement is performed, it will give a result whose difference from the mean is not much larger than σ.

So far, we have given a description of what we have called "experimental result" and we have explained how it ought to be expressed in mathematical terms. If, however, we want to use such a result to, say, falsify[3] a theory or model, then an interpretation of its meaning must be presented—meaning that can be expressed as the following: *the existence of a true value*[4] w_0 *for the quantity* w *is assumed*; the direct measurement of w_0 is, however, *impossible* due to (i) the fallibility of the measurer, (ii) the fallibility of the measuring instrument and (iii) environmental interference with the measured object. The "width" of the distribution is determined by these three factors combined. The only course of action the experimenter can take in this scenario is to extract from the distribution the best possible estimative of w_0 and, more importantly, to assess how good this estimate is. The way this is done is by putting the so called *central-limit theorem* (CLT) to use.

[1] It is preferred the set to have statistically independent elements. This means, ideally, repetitions must be performed by different measurers using distinct instruments ("distinct" meaning not the very same tool).

[2] "Real numbers" here is used in the mathematical sense.

[3] "Falsify" is here meant in the Popperian sense (to compare an experimental result with the value predicted by a theory in order to accept or reject it. See Popper (2002 [1934])).

[4] "True value" is the term used by statisticians to refer to a real number that represents the physical quantity in its true form; a number one would obtain by performing an error free, ideal measurement.

The CLT is represented in figure 1. It states that, *if the experiment is reproduced infinite times, each time yielding a mean* $\overline{w}_i$ *and a standard deviation* σ*, the distribution of the means* $\overline{w}_i$ *has a gaussian (or normal) form with mean identical to the true value* w_0 *and standard deviation identical to* $\sigma/\sqrt{N}$. The standard deviation of this gaussian distribution is called *uncertainty*. The CLT implies that the probability of a new experimental result $\overline{w}_i$, obtained by measuring N times the size w, to be within a certain interval from the true value w_0 can be calculated. That is because CLT yields the PDF, the gaussian curve, that describes the distribution of means. This allows the exact calculation of areas under the curve – of probabilities, that is.

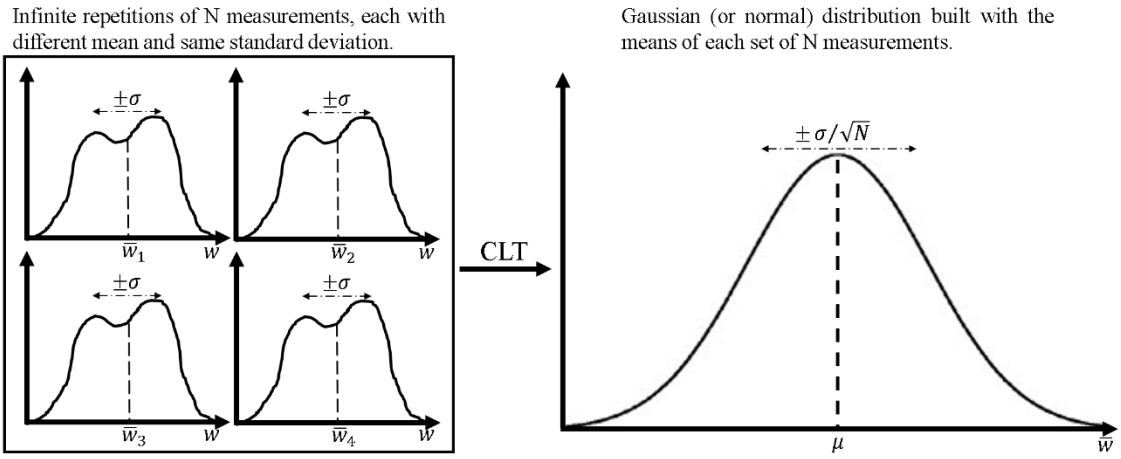

Figure 1:illustration of the central limit theorem. The distributions of the measurements do not need to be gaussian for the distribution of the means to be gaussian.

We close this section by making the following remark: falsification of theories and comparison between experimental results can only be expressed in terms of probability. An experiment attempting to falsify a theory can only yield conclusions such as "there is high (or low) probability such and such that theory is compatible with experimental results" and nothing of greater epistemological import. Probabilistic considerations, note, condition our very epistemic relation with experience.

3. Least squares method

The least squares method (LSM) is a method for measuring parameters of functions, or, equivalently, for measuring functions. Its importance can be summed up as follows: when one has data in one's hands, one can depict it in a graph that makes explicit important physical traits of such data—the velocity of an object in function of its position, say. Graphical analysis allows one to grasp important knowledge about the data depicted. There are, however, many ways functions can be drawn. More precisely, when one knows her data can be described by a certain kind of function, let's say a parabola, for instance, the question of which mathematical parameters best relate the measured physical quantities poses itself. This is, in a nutshell, what the LSM does for us: it helps us determine which set of

parameters of a chosen mathematical function best describes the behavior of the data—or, in simple words, how to best draw a graph.

In a first glance, LSM may appear to have nothing (or little) to do with the quantification of uncertainty of one single physical quantity, but there are at least three strong connections between them: (I) a single physical quantity can be thought of as the particular case of a function $f(x) = \text{constant}$; (II) parameters of functions must also have uncertainty associated to them; (III) historically, LSM and the uncertainty concept are closely related. Without further delay, let us see how the method works.

In the previous section we have seen that the usual measurement procedure of a physical quantity leads to a set of data which is interpreted as being not identical to but *distributed around* the true value of this quantity. Now, let us use this lingo in a situation where a function is to be measured, *i.e.*, the data is a set of pairs (x_i, y_i). x is here taken to denote the independent variable; y, the dependent one. Reproducing the previously presented argument, the parameters which determine the function are also taken to have true values. Consequently, for each true value the variable x can take, y will also have a true value. Given the fact that all we can do is measure fluctuations around true values, however, follows that each measured pair (x_i, y_i) is a frame of the fluctuation around the function. In this scenario, our job is to make the best possible estimate of the "true function" (true values of the function's parameters).

The first thing to do is to establish which function we are trying to measure, which function best describes the behavior of the data. Let's take the simple case of the equation of a straight line:

$$f(x) = ax + b. \tag{1}$$

To make it concrete, think of a body in uniform linear motion. In this case, x denotes time; $f(x)$ denotes the position (as a function of time, of course). These are the quantities that one will measure in this case: at the end of the day one will have a data set of measured pairs $(x, f(x))$. a and b, by their turn, denote velocity and initial position respectively (not directly measured). These are the parameters one would seek to indirectly quantify. Assuming the existence of true values for parameters a and b we conclude that, for each measurement x_i of x, there is a $f(x_i)$ representing the true value of the quantity y – let's call it μ_i. However, our measurement y_i is just a frame of the unavoidable fluctuation around μ_i. We call "error" and denote by E the difference between the measurement y_i and the true value μ_i.

$$E_i = y_i - \mu_i. \tag{2}$$

From a practical point a view, the error definition is not very useful, because we never know the true value μ_i. The closest we can get to quantifying the error is quantifying the difference between the measurement y_i and the estimated value of μ_i. This difference is called *residual*.

$$R_i = y_i - f(x_i) \tag{3}$$

We can also define, so to speak, a kind of total residual for the whole set of measurements. One convenient way of doing so is taking the sum over i of R_i^2 over the uncertainty of y_i:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2 \tag{4}$$

But why should we do so? Well, (4) allows us to define a criterion for estimating a and b: the best estimate for a and b is that which minimizes the value of χ^2. The way we defined χ^2 implies it must have an expected value which allows assessment of the quality of the estimates. The denominator in expression (4) works a statistical weight for each term of the sum: the more precise the measurement, the greater the weight it has on the sum.

Our final task is to find the expressions for a and b which minimize χ^2. Once χ^2 is defined, finding the needed expressions is just a simple calculus exercise. The value of a for which χ^2 is minimum is that which satisfies the equation:

$$\frac{\partial \chi^2}{\partial a} = 0 \tag{5}$$

The same is true for b:

$$\frac{\partial \chi^2}{\partial b} = 0 \tag{6}$$

Finally, we have the expressions for the estimative of a and b and they only depend on the measured (x_i, y_i) values.

$$a = \frac{\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} - \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} - \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}} \tag{6}$$

$$b = \frac{\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} - \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} - \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}} \tag{6}$$

The reader must have in mind the fact that the quality of the least square estimate must itself be assessed. A detailed presentation of how such an analysis works in practice would take us too far astray, however, and will not be done here. Suffices to note that an assessment of LSM is done by checking if the obtained value of χ^2 is probabilistically compatible with its PDF. Empowered by such tools, note, we know not only to assess how well the experiment was performed and how experimental results reflect on theory, but we can also guarantee such analysis is solid.

The problems displayed in sections two and three are, in fact, the same: *combining observations (elements of a data set) which are apparently analytically incompatible*.[5] The short introduction above given may hide the nontrivial character of such problem. Its complexity is made explicit by the acknowledgment of the long history behind its solution, a history that is briefly sketched in the next section. Focus will be given to the development of LSM. With such a sketch we hope to pave the way and give ground to the philosophical and epistemological considerations drawn in section 5.

4. A brief history of the conception of uncertainty

The problems of determining the value of a physical quantity from a set of measurements of this quantity and of determining parameters of functions from a set of measurements date both from at least 300 years BCE. For instance, between 500 and 300 years BCE, Babylonians developed mathematical tools, which required the estimation of parameters, for calculating the motion of some celestial bodies as a function of time. Unfortunately, no material remained for indicating how such estimates were done—it is only clear that they had to do estimations somehow in order to use their tools. Another example is reported by Manitius (1913): Hipparchus' endeavor, around 300 BCE, to determine whether or not the passage of the Sun through the same solstitial point is truly periodic. Hipparchus concluded the Sun does not pass the same point periodically by comparing his measurement with an error estimate of his creation. He determined the maximum variation in the duration of a year is ¾ of a day whilst his measurements' error could not be higher than ¼ of a day. He did not establish a way of calculating a representative value from a set of measurements, nor constructed a universal method for quantifying error. Neither mathematics nor concepts were ripe enough for doing so. His procedure did contain, however, at least implicitly, the idea of "fluctuations of measurements" due to errors.

A very interesting historic case in which the necessity of comparing experimental values has proved to be pressing is the so-called "trial of the Pyx", which was extensively studied by Stigler (1977). The trial of the Pyx is an event which occurs from time to time in Great Britain since, at least, 1248. Its purpose is to evaluate the quality of the coins produced by the Royal Mint. Even though details of the trial have changed through time, its general aspects can be summarized as follows: everyday, one coin out of a pre-established number of coins is taken from the Mint's production and stored in a box called "Pyx". After two or three years of repeated storage, the trial happens; the Pyx is opened and an assessment of the stored coins is performed in order to check if the coins meet predetermined standards regarding its weight and fineness. If they do not, the master of the Mint could face severe punishment. The trial, in a nutshell, is a straightforward case of estimating a population out of a sample; and we know it cannot be done without considering statistical fluctuations. What is particularly interesting here is the

[5] In section two, we have multiple measurements of the same object, which are not numerically equal to one another. In section three, we have multiple measurements forming what should be a straight line, but they cannot satisfy simultaneously any straight-line equation.

fact that, in fact, they did account for statistical fluctuations, even though in a rudimentary manner: a remedy (tolerance) was allowed; the coins had to be within this remedy. Stigler (1977) concluded that the remedy was very permissive with the master of the Mint. The remedy was too large; a skilled master could enrich greatly by pocketing a small fraction of silver and gold while still attending to the remedy. Ironically, the most prominent master of the Mint was Sir Isaac Newton, a mind no one would object to call "skilled". Newton served the position from 1699 to 1727. It is a historical fact that he became wealthy during his years in service of the Royal Mint, which raises the question whether or not he was taking advantage of the excessive tolerance. Actually, he faced charges regarding the fineness of the coins during the 1710 trial. However, de Villamil's (1931) and Craig's (1946) investigations lead us to believe Newton became wealthy as a result of his fair earnings and finance management.

Newton's "not guilty" verdict raises another question: why Newton did not defraud the Royal Mint? Was he following a moral compass or he simply did not realize the shortcoming in the trial's assessment? The answer is unclear, but Newton's last work – *The Chronology of Ancient Kingdoms Amended*, published *post mortem*, in 1728 – allows for speculations that he did know something about error theory. The mentioned work, a report to Princess Caroline, is an attempt to estimate the mean duration of reigns. Newton had at his disposal a table with mean reign durations of 12 kingdoms. This table had values ranging from 11.6 (Babylon) to 25.18 (Egypt) years. In the absence of statistical knowledge, we can speculate, one would state that the mean duration of a reign is something in the neighborhood of 18.4 ± 6.8 years, *i.e.*, the average between extremes ± half the difference between extremes; or the arithmetic mean ± half the difference between extremes; or just a straightforward "from 11.6 to 25.18". However, Newton's assertion was that kingdoms last "about eighteen to twenty years". Here is the astonishing part: the calculation, using Newton's table, of the arithmetic mean ± uncertainty, as presented in topic 1, leads to 19.1 ± 1.0 years! Saying that Newton had all error theory solved is certainly too much of an extrapolation. However, risking being accused of whiggish historicism notwithstanding, one can speculate that is very likely that Newton understood, at least intuitively, the inverse proportion between uncertainty and size of data set.

Let us now present the historical facts more directly connected to the development of tools for combining observations. Before late XVI century it was common procedure that, when one wanted to compare a result with a set of results, one would arbitrarily choose values within such a set to draw conclusions. Plackett (1958) states that Tycho Brahe, in the decade of the 1580s, appears to be the first to combine measurements in order to obtain a single value of a physical quantity. Brahe measured the right ascension[6] of the star α Arietis using different techniques and calculated the arithmetical mean between obtained values, in a clear attempt to remove systematic errors[7] from his results. However, the

[6] *α Arietis* is the brightest star in the northern zodiacal constellation of Aries. *Right Ascension* is, together with *declination*, a celestial coordinate for indicating a point on the celestial sphere.

[7] *Systematic error* is an error generated by an accuracy deficiency in an experiment. A clear example of systematic error is given by an uncalibrated instrument; which will cause an equal shift from the true value in all elements of a data set. However, unlike the uncalibrated instrument example, sometimes systematic errors cannot be avoided.

use of arithmetical mean as representative value of a data set became a considerable standard procedure only by the first half of the XVIII century (Plackett, 1958). By 1750, some astronomers had realized that combining observations making use of arithmetic mean could be advantageous somehow, but they would only combine measurements which were considered to have the same accuracy (Stigler, 1986: 16), *i.e.*, measurements performed exactly with same conditions (same measurer, time, space, instrument etc.). Possibly, Roger Cotes was the first one to express this idea objectively in a posthumously published work in 1722.

Leonhard Euler's study of Saturn and Jupiter shows clearly that, by mid-XVII century, combining observations was still not a well-established idea (Stigler, 1986: 25). In 1748, the Academy of Science in Paris announced a prize for the one who would provide the best explanation for inequalities observed in the orbit of Saturn and Jupiter. Over 50 years earlier Halley had observed that the former planet appeared to be retarding while the latter was accelerating; he also proposed the mutual attraction between the planets was the reason behind these inequalities. Euler, in his 1749's work, engaged in solving this problem. Assuming Saturn and Jupiter orbit the Sun elliptically and that the ellipses are not in the exact same plane, Euler came up with a fifteen parameter equation; seven of which were directly observable variables and the other eight were constants that could only be extracted from fitting. Euler then faced an embarrassment of riches: he had seventy-five sets of observations, *i.e.*, equations with experimental values, at his disposal, but only eight unknowns to find. He combined small sets of equations with similar coefficients, subtracting one from the other in an attempt to make those coefficients disappear. However, there were not enough observations with similar coefficients - leading him to a dead-end. Even though Euler managed to win the prize offered by the Academy of Science, he clearly failed to provide any meaningful way of combining observations. His failure catches the eye even more when contrasted to Tobias Mayer's successful work published just one year after.

As many great scientific endeavors of the eighteenth century, Mayer's work was also about astronomy and was closely related to technology and matters of state of the time (Stigler, 1986: 16). In 1714, seeking to improve the power of localizing ships on the sea, England established the "commissioners for the discovery of longitude at sea", an institution that offered prizes for the ones who would help find a way of determining longitude at the sea. In 1747, Mayer engaged in solving the commissioners' problem by studying the Moon - a task that led to a publication, in 1750, on the libration of the moon.[8] During the two preceding years, Mayer performed measurements of the position of some lunar features, which allowed him to infer characteristics of the lunar orbit. His observations were described by a linear equation with 3 measurable variables and 3 unknown constants. Therefore, he only needed three observations to solve the problem mathematically. His measurements, however, were

[8] The so called "dark side of the moon" is not, in fact, one-hundred percent dark. During a one-month period is possible to observe that the moon oscillates a little, showing us, during this time, almost sixty percent of its surface. This oscillatory movement is called "libration".

twenty-seven in total. Again, as in Euler's study of Saturn and Jupiter, an embarrassment of riches emerged; but Mayer's reaction to the problem was fundamentally different from Euler's. Mayer divided the twenty-seven observations in three groups of nine; then, he added the nine, turning the nine into one, and solved the three resulting linear equations. The fact that he understood the importance of not arbitrarily discarding experimental results together with his simplistic away of combining measurements is, by itself, impressive. His most remarkable achievement, however, was the way he grouped equations.

Mayer choose the groups based on the values of one of the unknowns. The first group had the biggest values of this unknown; the second group had the smallest values; and the third group had the "intermediate values". Curiously, Mayer mistakenly assigned the third group as having the biggest values of another constant. After finding the values of the unknowns, Mayer moved to analyzing the accuracy of his results. First, he defined error as "how far the quantity (...) can deviate from the true value". He stated clearly that considered the errors (he used the German word "*Fehler*") to be inversely proportional to the number of measurements. Then, he estimated the error and presented his final result in the form $x \pm error$. This allows us to conclude Mayer's arrangement of equations was not a mere accident. In his procedure, he assumes the existence of errors and true values; instead of trying to swipe the errors under the carpet by choosing only the convenient observations, as Euler tried to do, he separated the observations so as to consider the full extent of fluctuations. He took the philosophical position where increase in information implied increase in precision and built his procedure on it.

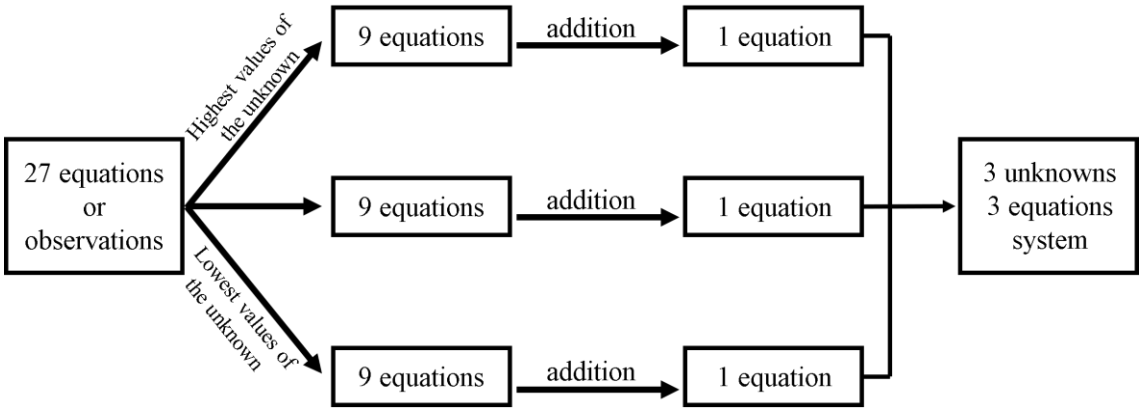

Figure 2: scheme of Mayer's procedure of combining equations in his study on the libration of the moon.

It did not take long for Mayer's method to become fairly accepted. It was widely used during the second half of the eighteenth century, before LSM's birth. Despite the success due to its easy applicability and easy handling, the method lacked a clear theoretical foundation. For instance, it stated nothing about the quality of the results, *i.e.*, it did not answer the question "is the obtained result the closest to the true value one can get?". Boscovich, in 1760, was the first to express the combination of observations problem in terms of general statistical principles. Just as Euler and Mayer, Boscovich was also working on astronomical observations. By that time, there was still doubt about the precise shape

of the earth[9] and Boscovich was analyzing data related to this problem (Stigler, 1986: 39). What he proposed in his analysis is a primitive version of χ^2. He proposed that (I) the sum of all residuals r_i (he used the word "correction") must be zero and; (II) the sum of the absolute values of the residuals must be minimum.[10]

$$\sum_{i=0}^{N} r_i = 0\,; \tag{7}$$

$$\sum_{i=0}^{N} |r_i| is\ minimum. \tag{8}$$

Boscovich did not realize he could differentiate his expressions as we did in section 3. Instead, he built a very interesting and overly complicated geometrical solution. For the lack of practicality, his solution would probably not reach very far regarding the impact on his peers. Even Boscovich himself did not go back to use his method again. But the sharp eyes of Laplace noted the potential of Boscovich's idea.

In 1789, Laplace, also while working on the shape of the Earth problem, having perceived the excessive complication in Boscovich's geometrical solution, published an analytical version of Boscovich's method. Ten years later, in 1799, Laplace moved further and added one important ingredient to Boscovich's axioms. He gave statistical weight to each term in expression (7), understanding that the higher the precision of the measurement, the higher should be its weight (Stigler, 1986: 50). Even though he did not quantify the weight in a manner we would approve today, the importance of the step he took forward is indisputable.

LSM's first appearance was in an appendix of Legendre's *memoir on the determination of cometary orbits* (1805). It is clear that, at this point, after so many incremental advances, the stage was set for the birth of LSM. All the conceptual grounds had already been built. The only thing still missing was someone to realize that the minimization of the sum of the square of the errors was a better and easier to handle criterion than Boscovich's sum of the absolute values; and that is exactly what Legendre did in his memoir.

5. Physics and philosophy join hands

Three general, but nonetheless important, lessons can be drawn from the preceding sections. Part of the physicist's armory when dealing with data analysis was described in sections 2 and 3. The

[9] By that time, it was a consensus that the Earth was not exactly sphere shaped, but a sphere with flattened fractions. The discussion was limited to where the flattened parts were: at the poles or at the equator.

[10] These criteria appear to be very similar to the χ^2 presented in section 3, but they present some flaws that do not remain in χ^2 (Nievergelt, 2000).

road traveled for some of such arsenal to be forged, whigguishly outlined in section 4. This historical sketch, its briefness notwithstanding, is indisputable evidence against the view of scientific discovery that takes important scientific results to be "found" by one particularly gifted person in a final frenzy of intellectual effort: from Hipparchus to Legendre there is a time lapse of millenia. Although the err of such a conception is obvious to most philosophers, well-acquainted as they are with history of science, this is not the case of scientists—it is, in fact, a position endorsed by most of them. Such tales about how theories are forged are found in many physics textbooks widely used in courses of basic physics—notorious examples are (Nussenzveig, 1998: 182) and (Halliday *et al*., 2007: 1253-1254) at introducing Einstein's formulation of special relativity. At reading their presentation one falls under the impression the relativity of simultaneity has reached Einstein's thought as suddenly as a stroke—sweeping under the carpet all intellectual historical context that has proven to be *sine qua non* condition for such a scientific revolution to be in order. Physics students sleep over such books and it is fair to assume most of them believe in their content. How much such views of scientific discovery influences one's own view of one's work and ability to advance knowledge is a question we here leave open, but it is reasonable to assume it is nonnegligible.

Philosopher of science Neurath (1921) has once compared science to a boat that floats in the open sea. The scientists and the philosopher, he wrote, are on the same boat. Examples were given in section 4 in which the solution advanced towards a problem relied upon a preconceived idea of how the solution would look like—preconceptions demonstrably philosophical in character. This is clear for instance when one contrasts Euler's and Mayer's attempts to combine observations: the former, his mathematical brilliance notwithstanding, could not face the matter with the appropriate philosophical stand—the stand that has led the latter to the right solution. In effect, this is one of the senses in which the authors want to be understood when advocating the interdependence of science and philosophy: just as the mathematician yields to the physicist a box in which she can find several tools she may put to use at will to the purposes she thinks better fit, in a similar vein philosophers are experts in uncovering details of problems and advancing metaphysical and epistemological tools to approach them that can be used profitably by scientists—on the condition they are aware such tools are at their disposal. Sometimes the situation is such tools needed had not even been crafted, however—as in the case above mentioned—, and scientific and philosophical thinking must merge and be put to work at once for the appropriate tool to be forged. This is one way of stating the authors' position that science and philosophy are, at the end of the day, sides of one same coin. What matters, when one seeks to find solution to problems such as those faced by Newton, Euler, and Mayer, is originality of approach—and that's when the philosopher's toolbox kicks in and she and the scientist may work jointly to advance clever solutions to relevant questions.

In closing, we make the following remark about statistical methods in science and their implications: a theory's final fate lies always at the hands of the experimentalist—Nature, *experience*, is always the ultimate judge. Sections 2 and 3 make clear that all epistemic contact we have with

experience is intrinsically probabilistic in character; the ways with which science uncovers the workings of Nature are limited by the constraints of probability theory. The philosophical resonance of this indisputable empirical fact is deep and far-reaching and will be pursued in other paper.

References

CRAIG, J. Newton at the Mint. Cambridge: Cambridge University Press, 1946.
DE VILAMIL, R. Newton: The Man. London: Knox, 1931.
EARMAN, J. Bayes or Bust? Cambridge, US: MIT Press, 1992.
HALLIDAY, D.; RESNICK, R.; WALKER, J. Fundamentals of Physics. New Jersey, US: John Wiley & Sons, 2007.
HELENE, O.; TSAI, S.; TEIXEIRA, R. What is a Measurement? Revista Brasileira de Ensino em Física, vol. 13, p. 12-29, 1991.
MANITIUS, K. Des Claudius Ptolemäus Handbuch der Astronomie. Leipzig: Teubner, 1913.
NEURATH, O. Anti-Spengler. Vienna: Callwey, 1921.
NIEVERGELT, Y. A Tutorial History of Least Squares with Applications to Astronomy and Geodesy. Journal of Computational and Applied Mathematics, v. 121, p. 37-72, 2000.
NUSSENZVEIG, H. Curso de Física Básica – vol. 4. São Paulo, Brazil: Blucher, 1998.
PLACKETT, R. Studies in the History of Probability and Statistics VII: The Principle of the Arithmetic Mean. Biometrika, vol. 45, n. 1/2, p. 130-135, 1958.
PLACKETT, R. Studies in the History of Probability and Statistics XXIX: The Discovery of the Method of Least Squares. Biometrika, vol. 59, n. 2, p. 239-251, 1972.
POPPER, K. The Logic of Scientific Discovery. Abingdon-on-Thames: Routledge, 2002.
SACHKOV, Y. Probability in Classical and Quantum Mechanics. Studies in Logic and the Foundations of Mathematics, vol. 104, p. 441-447, 1982.
SKLAR, L. Probability as a Theoretical Concept. Synthese, vol. 40, p. 409-414, 1979.
STIGLER, S. The History of Statistics: The Measurement of Uncertainty before 1900. Cambridge, Massachusetts, Us: Harvard University Press, 1986.
STIGLER, S. Gauss and the Invention of Least Squares. The Annals of Statistics, vol. 9, n. 3, p. 465-474, 1981.
STIGLER, S. Eight Centuries of Sampling Inspection: The Trial of the Pyx. Journal of the American Statistical Association, vol. 72, n. 359, p. 493-500, 1977.
TAYLOR, J. An Introduction to Error Analysis. California, US: University Science Books, 1997.
VON WEIZSÄCKER, C. Probability and Quantum Mechanics. The British Journal for the Philosophy of Science, vol. 24, p. 321-337, 1973.
VUOLO, J. Fundamentos da Teoria de Erros. São Paulo, Brazil: Blucher, 1996.