

# The Gauss' Bayes Factor

Giulio D'Agostini

Università “La Sapienza” and INFN, Roma, Italia

(giulio.dagostini@roma1.infn.it, <http://www.roma1.infn.it/~dagos>)

## Abstract

In *Theoria motus corporum coelestium in sectionibus conicis solem ambientium* Gauss presents, as a theorem and with emphasis, the rule to update the ratio of probabilities of complementary hypotheses, in the light of an observed event which could be due to either of them. Although he focused on *a priori* equally probable hypotheses, in order to solve the problem on which he was interested in, the theorem can be easily extended to the general case. But, curiously, I have not been able to find references to his result in the literature.

“I play with a gentleman whom I do not know.  
 He has dealt ten times,  
 and he has turned the king up six times.  
 What is the chance that he is a sharper?  
 This is a problem in the probability of causes.  
 It may be said that it is the essential problem  
 of the experimental method.”  
 (H. Poincaré)

## 1 Introduction

As it is becoming rather well known, the only sound way to solve what Poincaré called “*the essential problem of the experimental method*” is to tackle it using probability theory, as it should be rather obvious for “*a problem in the probability of causes*”. The mathematical tool to perform what is also known as ‘probability inversion’ is called *Bayes rule* (or theorem), although due to Laplace, at least in one of the most common formulations:<sup>1</sup>

$$P(C_i | E, I) = \frac{P(E | C_i, I) \cdot P(C_i | I)}{\sum_k P(E | C_k, I) \cdot P(C_k | I)}, \quad (1)$$

---

<sup>1</sup>“The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause [given that event]. The probability of the existence of any one of these causes [given the event] is thus a fraction

where  $E$  is the observed event and  $C_i$  are its possible causes, forming a complete class (i.e. exhaustive and mutually exclusive). ‘ $I$ ’ stands for the background state of information, on which all probability evaluations *do* depend (‘ $I$ ’ is often implicit, as it will be later in this paper, but it is important to remember of its existence).

Considering also an alternative cause  $C_j$ , the ratio of the two *posterior probabilities*, that is how the two hypotheses are re-ranked in *degree of belief*, in the light of the observation  $E$ , is given by

$$\frac{P(C_i | E, I)}{P(C_j | E, I)} = \frac{P(E | C_i, I)}{P(E | C_j, I)} \times \frac{P(C_i | I)}{P(C_j | I)}, \quad (2)$$

in which we have factorized the r.h. side into the *initial* ratio of probabilities of the two causes (second term) and the updating factor

$$\frac{P(E | C_i, I)}{P(E | C_j, I)}, \quad (3)$$

known as *Bayes factor*, or ‘likelihood ratio’.<sup>2</sup> The advantage of Eq. (2) with respect to Eq. (1) is that it highlights the two contributions to the *posterior* ratio of the hypothesis of interest: the *prior* probabilities of the ‘hypotheses’, on which there could be a large variety of opinions; the ratio of the probabilities of the observed event, under the assumption to each hypothesis of interest, which can *often* be rather *intersubjective*, in the sense that there is usually a larger, or unanimous consensus, if the conditions under they have been evaluated (‘ $I$ ’) are clearly stated and shared (and in critical cases we have just to rely on the well argued and documented opinion of experts.<sup>3</sup>)

---

*whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes. If the various causes are not equally probable a priori, it is necessary, instead of the probability of the event given each cause, to use the product of this probability and the possibility of the cause itself.”*[1]

<sup>2</sup>The alternative name *likelihood ratio* is preferred in some communities of researchers because numerator and denominator of Eq. (3) are called *likelihood* by statisticians. My preference to ‘Bayes factor’ (or even *Bayes-Turing Factor* [2, 3, 4]) is due to the fact that, since in the common parlance ‘likelihood’ and ‘probability’ are in practice equivalent, ‘likelihood ratio’ tends to generate confusion as it were the ratio of the probabilities of the hypotheses of interest (and the value that maximizes the ‘likelihood function’ tends to be considered *by itself* the most probable value).

<sup>3</sup>For example the European Network of Forensic Science Institutes strongly recommends [5] forensic scientists to report the ‘likelihood ratio’ of the findings in the light of the hypothesis of the prosecutor and the hypothesis of the defense, abstaining to assess which hypothesis they consider more probable, task left to the judicial system (but then I have strong worries, shared by other researchers, about the ability of the members of judicial system of making the proper use of such a quantitative information!). To those interested on the details of how this *Guideline* can be turned into practice, a *Coursera* offered by the University of Lausanne is recommended [6]

Recently, going after years through the third section of the second ‘book’ of Gauss’ *Theoria motus corporum coelestium in sectionibus conicis solem ambientum* [7, 8], of which I had read with the due care only the part in which the *Prince Mathematicorum* derives in his peculiar way what is presently known as the Gaussian (or ‘normal’) error function, I have realized that Gauss had also illustrated, a few pages before, a theorem on how to update the probability ratio of two alternative hypotheses, based on experimental observations. Indeed the theorem is not exactly Eq.(2), because it is only formulated for the case in which  $P(C_i | I)$  and  $P(C_j | I)$  are equal, but the reasoning Gauss had setup would have led naturally to the general case. It seems that he focused into the sub-case of *a priori* equally likely hypotheses just because he had to apply his result to a problem in which he consider the values to be inferred *a priori* equally likely (“*valorum harum incognitarum ante illa observationes aequae probabilia fuisse*”).

But let us proceed in order.

## 2 Probability of observations vs probability of the values of physical quantities

The third section of ‘book 2’ of the Gauss’ tome [7] is dedicated to “*the determination of an orbit satisfying as nearly as possible any number of observations whatever*”.<sup>4</sup> After ‘articles’<sup>5</sup> 172-174, which introduce the specific problem of evaluating the elements of an orbit from the measurements of geocentric quantities related to those elements, with article 175 Gauss *ascends*<sup>6</sup> to methodological issues of general interest for the Sciences:

*“let us leave our special problem, and enter upon a very general discussion and one of the most fruitful in every application of the calculus to the natural philosophy.”*

The general problem is how to determine the  $\mu$  unknown quantities  $p, q, r, s$ , etc. (e.g. the elements of the orbit of a planet or a comet) and evaluate the functions  $V_i$  of these variables from  $\nu$  measurements  $V_{m_i}$  (e.g. the geocentric quantities of

---

<sup>4</sup>All English quotes are taken from the C.H. Davis translation [8].

<sup>5</sup>This publication is divided into two ‘books’, each of them subdivided in four ‘sections’. Then the entire text is divided in *numeri* (translated into ‘articles’ by Davies [8]) running through the ‘books’. In particular, Section 3 of Book 2, consisting of 22 printed pages in the original Latin edition, contains ‘articles’ 172 to 189.

<sup>6</sup>“... *ad disquisitionem generassimam in omni calculi ad philosophiam naturalem applicatione fecundissima ascendemus.*”

that celestial body measured at different times):<sup>7</sup>

$$V_i(p, q, r, s, \dots) \xrightarrow{\text{measured as}} V_{m_i}, \quad (4)$$

or, indicating the set of unknown quantities by  $\boldsymbol{\theta}$ , that is  $\boldsymbol{\theta} = \{p, q, r, s, \dots\}$ , we can rewrite Eq. (4) as

$$V_i(\boldsymbol{\theta}) \xrightarrow{\text{measured as}} V_{m_i}. \quad (5)$$

The most interesting case, Gauss explains, is when  $\nu > \mu$ . Being over-determined, this case has a solution only if  $V_{m_i}$  are affected by experimental errors, described by the probability density function<sup>8</sup> (pdf)  $\varphi$ , that in article 177 will come out to be the well known Gaussian function.<sup>9</sup> Therefore,<sup>10</sup>

“Supposing, therefore, any determinate system of the values of the quantities  $p, q, r, s$ , etc., the probability that the observation would give for  $V$  the value  $M$  will be expressed by  $\varphi(M - V)$ , substituting in  $V$  for  $p, q, r, s$ , etc., their values; in the same manner  $\varphi(M' - V')$ ,  $\varphi(M'' - V'')$ , etc, will express the probability that observation would give the values  $M', M''$ , etc. of the functions  $V', V''$ , etc. Wherefore, since we are authorized to regard all observations as event independent of each other, the product

$$\varphi(M - V) \varphi(M' - V') \varphi(M'' - V'') \text{ etc,} = \Omega \quad (\text{G1})$$

will express the expectation or probability that all those values will result together from observation.”

What Gauss calls  $\Omega$  is thus the *joint pdf* of the differences  $V_{m_i} - V_i$  given a precise set of values for the physical quantities of interest, which we would rewrite as

$$f(\mathbf{V}_m - \mathbf{V} | \boldsymbol{\theta}) = \prod_i \varphi(V_{m_i} - V_i | \boldsymbol{\theta}) \quad (6)$$

---

<sup>7</sup>For the reader's convenience (hopefully) the functions are called here, except when they appear in quotes,  $V_1, V_2, V_3$ , etc., and the measured values  $V_{m_1}, V_{m_2}, V_{m_3}$ , etc., while Gauss uses  $V, V', V'' \dots$  and  $M, M', M'' \dots$ , respectively.

<sup>8</sup>Note how Gauss simply speaks of ‘probabilities’, obviously meaning *probability density functions*, as clear from the use he makes of them: “the probability to be assigned to each error  $\Delta$  will be expressed by a function of  $\Delta$  that we shall denote as  $\varphi\Delta$ ” – a few lines later it is clear that Gauss had in mind a ‘pdf’ since, when he wrote “the probability generally, that the error lies between  $D$  and  $D'$ , will be given by the integral  $\int \varphi\Delta d\Delta$  extended from  $\Delta = D$  to  $\Delta = D'$ ”. [Note also that in the case a function had only one argument, parentheses were not used. Therefore  $\varphi\Delta$  stands for  $\varphi(\Delta)$ .]

<sup>9</sup>For an account of Gauss' derivation in modern notation see Sec. 6.12 of Ref. [9] (some intermediate steps needed to reach the solution are sketched in [http://www.roma1.infn.it/~dagos/history/Gauss\\_Gaussian.pdf](http://www.roma1.infn.it/~dagos/history/Gauss_Gaussian.pdf)).

<sup>10</sup>As clarified in footnote 8, it is clear that in the following quotes the generic term “probability” stands for *probability density function*.

where  $\mathbf{V}_m$  and  $\mathbf{V}$  stand for the set of observations and of functions.<sup>11</sup>

Article 175 ends so with the expression of the joint probability of the observations given any set of values of the quantities of interest, that is a problem in *direct probabilities*:

$$\boldsymbol{\theta} \xrightarrow{\text{deterministic link}} \mathbf{V} \xrightarrow{\text{probabilistic link}} \mathbf{V}_m$$

Article 176 begins with what we could call nowadays a ‘Bayesian manifesto’:

*“Now in the same manner as, when any determinate values whatever of the unknown quantities being taken, a determinate probability corresponds, previous to observations, to any system of values of the functions  $V, V', V''$ , etc; so, inversely, after determinate values of the functions have resulted from observation, a determinate probability will belong to every system of values of the unknown quantities, from which the values of the functions could possibly have resulted.”*

That is, in our notation, as when we assume “*determinate values*” of the physical quantities we are interested in the joint pdf of the values that will be observed,

$$f(\mathbf{V}_m - \mathbf{V} | \boldsymbol{\theta}), \quad (7)$$

similarly, once the observations have been made, we are interested in the joint pdf of the values of the physical quantities,

$$f(\boldsymbol{\theta} | \mathbf{V}_m - \mathbf{V}). \quad (8)$$

The question is now how to go from Eq. (7) to Eq. (8), reasoning “inversely”.

### 3 Updating the probabilities of hypotheses

We are finally at the core of the problem. Let Gauss speak:

*“For, evidently, those systems will be regarded as the more probable in which the greater expectation had existed of the event which actually occurred. The estimation of this probability rests upon the following theorem:*

*If, any hypothesis  $H$  being made, the probability of any determinate*

---

<sup>11</sup>The reason why in the argument appears the differences and not simply the observed values is that for Gauss  $\varphi()$  was the error function, i.e. ‘probability density function’ of the errors (see also footnote 8). Since, later in ‘article’ 177, the function  $\varphi()$  will become the ‘Gaussian’ error function, we could rewrite directly the joint pdf of the observations in modern notation as

$$f(\mathbf{V}_m | \boldsymbol{\theta}) = \prod_i \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(V_{m_i} - V_i(\boldsymbol{\theta}))^2}{2\sigma^2} \right].$$

tarum, e quibus illi demanare potuerunt, probabilitas determinata redundabit: manifesto enim systemata ea pro magis probabilibus habenda erunt, in quibus euentus eius qui prodiit exspectatio maior affuerat. Huiusce probabilitatis aestimatio sequenti theoremati innititur:

*Si posita hypothesi aliqua H probabilitas alicuius euentus determinati E est = h, posita autem hypothesi alia H' illam excludente et per se aeque probabili eiusdem euentus probabilitas est = h': tum dico, quando euentus E reuera apparuerit, probabilitatem, quod H fuerit vera hypothesis, fore ad probabilitatem, quod H' fuerit hypothesis vera, vt h ad h'.*

Figure 1: Extract of *Theoria motus corporum...* [7] in which Gauss enunciates his theorem on how to update probability ratios of incompatible hypotheses in the light of an experimental observation. Note “**tum dico**” (“than I say”).

*event E is h, and if, another hypothesis H' being made excluding the former and equally probable in itself, the probability of the same event is h': then I say, when the event E has actually occurred, that the probability that H was the true hypothesis, is to the probability that H' was the true hypothesis, as h to h'.*” (Italic original, also put in evidence in the text as a quote – see Fig. 1.)

In modern notation:

$$\begin{aligned} P(E | H) &= h \\ P(E | H') &= h' \\ \frac{P(H | E)}{P(H' | E)} &= \frac{P(E | H)}{P(E | H')}, \quad \text{if } P_0(H) = P_0(H'). \end{aligned} \quad (9)$$

There are no doubts that Gauss presents this result as original (“*then I say*”, in Latin *tum dico*), although it might be curious that it did not refer to results by Laplace, who had been writing on *probabilities of causes* more than thirty years before<sup>12</sup> [10]. (For comparison, a few pages later, in article 177, Gauss acknowledges Laplace for having calculated the integral needed to normalize the ‘Gaussian’ distribution.) It is also curious the fact that Gauss starts saying that “*evidently, those systems will be regarded as the more probable in which the greater expectation had existed of the event which actually occurred*”, considering thus “*evident*” what is presently known as ‘maximum likelihood principle’, but

<sup>12</sup>The historian of statistics Stephen Stigler refers to Laplace’s 1774 *Mémoire* as “*arguably the most influential article this field [mathematical statistics<sup>(\*)</sup>] to appear before 1800, being the first widely read presentation of inverse probability and its application to both binomial and location parameter estimation.*” [11] (note that in this reference there is no mention to Gauss). (\* As far as I know, neither Gauss nor Laplace were using the word ‘statistics’, but they were talking about probability.

Ad quod demonstrandum supponamus, per distinctionem omnium circumstantiarum, a quibus pendet, num  $H$  aut  $H'$  aut alia hypothesis locum habeat, vtrum euentus  $E$  an alius emergere debeat, formari systema quoddam casuum diuersorum, qui singuli per se (i. e. quandiu incertum est, vtrum euentus  $E$  an alius proditurus sit) tamquam aequae probabiles considerandi sint, hosque casus ita distribui,

vt inter ipsos reperiantur	vbi locum habere debet hypothesis	cum modificationibus talibus vt prodire debeat euentus
$m$	$H$	$E$
$n$	$H$	ab $E$ diuersus
$m'$	$H'$	$E$
$n'$	$H'$	ab $E$ diuersus
$m^p$	ab $H$ et $H'$ diuersa	$E$
$n^q$	ab $H$ et $H'$ diuersa	ab $E$ diuersus

Figure 2: Partition of the space of possibilities as it appears in the original work of Gauss [7]. The English translations of the three columns are [8]: “that among them may be found”; “in which should be assumed the hypothesis”; “in such a mode as would give occasion to the event”. Then: “ab  $E$  diuersus” = “different from  $E$ ”; “ab  $H$  et  $H'$  diuersa” = “different from  $H$  and  $H'$ ”.

then taking care of proving it as a theorem (under the well stated assumption of initially equally probable hypotheses).

The reasoning upon which the theorem is proved is based on an inventory of equiprobable cases. This might seem to limit the application to situations in which this inventory is in practice feasible, like in games of cards and of dice. Instead, this was the way of reasoning of those times to partition the space of possibilities, as it is clear from the use that Gauss makes of his result, certainly not limited to simple games. Figure 2 shows the original version of such a partition. The six numbers of the first column, normalized to their sum, provide the

following probabilities:

$$\begin{aligned}
P(E \cap H) &= \frac{m}{m + n + m' + n' + m'' + n''} \\
P(\bar{E} \cap H) &= \frac{n}{m + n + m' + n' + m'' + n''} \\
P(E \cap H') &= \frac{m'}{m + n + m' + n' + m'' + n''} \\
P(\bar{E} \cap H') &= \frac{n'}{m + n + m' + n' + m'' + n''} \\
P(E \cap \overline{H \cup H'}) &= \frac{m''}{m + n + m' + n' + m'' + n''} \\
P(\bar{E} \cap \overline{H \cup H'}) &= \frac{n''}{m + n + m' + n' + m'' + n''}
\end{aligned}$$

The probabilities which enter the proof are those of the  $H$  and  $H'$

$$P(H) = \frac{m + n}{m + n + m' + n' + m'' + n''} \quad (10)$$

$$P(H') = \frac{m' + n'}{m + n + m' + n' + m'' + n''} \quad (11)$$

and those of the event  $E$  given either hypothesis:

$$P(E | H) = \frac{m}{m + n} = h \quad (12)$$

$$P(E | H') = \frac{m'}{m' + n'} = h' \quad (13)$$

The probability of  $H$  is modified by the observation of  $E$  observing that, with reference to Eqs. (10) and (11),

*“after the event is known, when the cases  $n$ ,  $n'$ ,  $n''$  disappear from the number of possible cases, the probabilities of the same hypothesis will be*

$$\frac{m}{m + m' + m''};$$

*in the same way the probability of the hypothesis  $H'$  before and after the event, respectively, will be expressed by*

$$\frac{m' + n'}{m + n + m' + n' + m'' + n''} \quad \text{and} \quad \frac{m'}{m + m' + m''} :$$



since, therefore, the same probability is assumed for the hypotheses  $H$  and  $H'$  before the event is known, we shall have

$$m+n = m'+n', \quad (\text{G2})$$

hence the truth of the theorem is readily inferred.”

That is, in our notation,

$$P(H | E) = \frac{m}{m + m' + m''}$$

$$P(H' | E) = \frac{m'}{m + m' + m''},$$

from which

$$\frac{P(H | E)}{P(H' | E)} = \frac{m}{m'}.$$

Using then Eqs. (12) and (13), yielding  $m = (m + n) \cdot P(E | H)$  and  $m' = (m' + n') \cdot P(E | H')$ , we obtain

$$\frac{P(H | E)}{P(H' | E)} = \frac{P(E | H) \cdot (m + n)}{P(E | H') \cdot (m' + n')} \quad (14)$$

Applying finally the condition (G2), theorem (9) is proved.

In reality, it is easy to see that, being

$$\frac{m + n}{m' + n'} = \frac{P(H)}{P(H')},$$

Eq. (14) contains the most general case

$$\frac{P(H | E)}{P(H' | E)} = \frac{P(E | H)}{P(E | H')} \cdot \frac{P(H)}{P(H')}.$$

But Gauss contented himself with the sub-case of initially probable hypotheses. Why? The reason is most likely that he focused on the inference of the unknown values of the physical quantities of interest, that he assumed *a priori* equally likely, a very reasonable assumption for this kind of inferences, if we compare the prior knowledge with the information provided by observations (see e.g. Ref. [9]).

## 4 Application to the inference of unknown values of physical quantities

In fact, immediately after the proof of *his* theorem, Gauss continues:

“Now, so far as we suppose that no other data exist for the determination of the unknown quantities besides the observations  $V = M$ ,  $V' = M'$ ,  $V'' = M''$  etc., and, therefore, that all systems of values of these unknown quantities were equally probable previous to the observations, the probabilities, evidently, of any determinate system subsequent to the observations will be proportional to  $\Omega$ . This is to be understood to mean that the probability that the values of the unknown quantities lie between the infinitely near limits  $p$  and  $p + dp$ ,  $q$  and  $q + dq$ ,  $r$  and  $r + dr$ ,  $s$  and  $s + ds$ , etc. respectively, is expressed by

$$\lambda \Omega dp dq dr ds \cdots, \text{ etc.}, \quad (\text{G3})$$

where the quantity  $\lambda$  will be a constant quantity independent of  $p$ ,  $q$ ,  $r$ ,  $s$ , etc.: and, indeed,  $1/\lambda$  will, evidently, be the value of the integral of order  $\nu$ ,

$$\int^{\nu} \Omega dp dq dr ds \cdots, \text{ etc.}, \quad (\text{G4})$$

for each of the variables  $p$ ,  $q$ ,  $r$ ,  $s$ , etc, extended from the value  $-\infty$  to the value  $+\infty$ .”

As we can see, it is well stated the assumption of ‘flat priors’, as we use to say nowadays (with the original words of Gauss, in Latin: “*valorum harum incognitarum ante illa observationes aequae probabilia fuisse*”).<sup>13</sup>

It is, instead, less clear how he uses the result of his theorem (the quote at the beginning of this section follows immediately the end of the proof of the theorem, with no single word in between). The implicit intermediate step is

$$P(H | E) \propto P(E | H), \quad (15)$$

extended to set of continuous uncertain values (‘uncertain vector’)  $\boldsymbol{\theta}$  as

$$P(\boldsymbol{\theta} | \text{data}) \propto P(\text{data} | \boldsymbol{\theta}). \quad (16)$$

Then, remembering that  $\Omega$  was the joint pdf of the observations [see Eq. (G1)], which we have rewritten in more compact notation as Eq. (6), we have

$$f(\boldsymbol{\theta} | \mathbf{V}_m - \mathbf{V}) \propto f(\mathbf{V}_m - \mathbf{V} | \boldsymbol{\theta})$$

or

$$f(\boldsymbol{\theta} | \mathbf{V}_m - \mathbf{V}) = \lambda \cdot f(\mathbf{V}_m - \mathbf{V} | \boldsymbol{\theta}),$$

---

<sup>13</sup>It is clear that what is unknown are the numeric values of the quantities and not the ‘quantities’ themselves, at it could seem from the English translation, because in that case there would be little to infer.

being  $\lambda$  just the normalization constant, i.e.<sup>14</sup>

$$\frac{1}{\lambda} = \int_{\mathbb{R}^{\nu}} f(\mathbf{V}_m - \mathbf{V} | \boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (17)$$

## 5 Conclusions

Reading Gauss' work, there are no doubts that the *Prince Mathematicorum* had clear ideas on how to tackle inverse probability problems, i.e. what goes presently under the name *Bayesian inference*. In particular, he presented as original what is now called Bayes factor, i.e the factor to update the *odds* in favor of an hypothesis with respect to the alternative one, in the light of a new observation. However, it is curious that, as far as I could find, this result is not acknowledged in the current literature. For example, his name appears only once in the Sharon McGrayne rather comprehensive book on the history of Bayesian reasoning[13], as being cited by Enrico Fermi, who was teaching his students data analysis methods derived from *his* Bayes' theorem.<sup>15</sup>

At this point a long discussion could follow on the question if Gauss could be classified as a *Bayesian* and why, later on in his book, he did not proceed applying consistently the probabilistic reasoning he had setup, getting the joint probability distribution of the values of the orbital elements given the observed geocentric measurements, but he derived, instead, the *least square* method to get (relatively) simple formulae for the *most probable values* (this aim was clearly stated). And all this in the same text, just a few pages after, and not in a later stage of his life.

Well, I am not an historian, and therefore I can only state my impressions based on a limited amount of reading. Gauss appears in the section of the book upon which this modest note is based not only as the genius he is famous to be, but also a very practical scientist going straight to his goals. Trying to set a multi-dimensional inference to write down the joint pdf of parameters of a non-linear problem and exploiting it at best, something that we can do nowadays, thanks to unprecedented computing power and novel mathematical methods, would have just been a waste of time two centuries ago. We have also seen that he didn't even care to state the general rule to update probability ratios, which would have required just a couple of lines of text, because he had in mind a problem for which the priors were reasonable 'flat'. Moreover, he was also well aware of

---

<sup>14</sup>Note: the reason of using in the formulae  $\mathbf{V}_m - \mathbf{V}$ , instead than just  $\mathbf{V}_m$  is simply due to the way Gauss wrote the error function, but, obviously, this function could be redefined and  $\mathbf{V}$  would disappear from the above equations, then getting for example  $f(\boldsymbol{\theta} | \mathbf{V}_m) \propto f(\mathbf{V}_m | \boldsymbol{\theta})$ , as we would write it nowadays (see also footnote 11).

<sup>15</sup>Indeed Ref. [13] cites Ref. [14], writing which I had realized that Gauss was using a 'Bayesian reasoning', but I had at that time completely skipped the 'details' in which he derived, as a theorem, the rule to update the ratio of probabilities of hypotheses, subject of this paper.

the practical meaning and limits of the mathematical functions, as when, later in the same section, he commented in ‘article’ 177 on the “defect” of *his error function*, because “*the function just found cannot, it is true, express rigorously the probabilities of the errors*”. Indeed, the ‘error function’  $\varphi()$  was not specified up to the end of ‘article’ 176. Only in the following article he showed that a good candidate for it was, under well stated conditions, ... the Gaussian, a function having the “defect” of contemplating values ranging from minus infinity to plus infinity. Then other interesting articles follow,<sup>16</sup> but I don’t want to spoil you the pleasure of the reading.<sup>17</sup>

Finally, someone might be intrigued about what Gauss meant by *probability*. “*Probabilitas*”. What else?<sup>18</sup>

## References

- [1] P.S. Laplace, *Essai philosophique sur les probabilités*, 1814, <http://books.google.it/books?id=JrEWAAAAQAAJ> (English quotes in this paper are taken from A.I. Dale’s translation, Springer-Verlag, 1995).
- [2] I.J. Good, *A list of properties of Bayes-Turing factors*, unclassified by NSA in 2011, <https://www.nsa.gov/news-features/declassified-documents/tech-journals/assets/files/list-of-properties.pdf>.
- [3] D.L. Banks, *A conversation with I.J. Good*, *Statistical Science*, Vol. 11, No 1, 1-11, 1996, [https://projecteuclid.org/download/pdf\\_1/euclid.ss/1032209661](https://projecteuclid.org/download/pdf_1/euclid.ss/1032209661).
- [4] G. D’Agostini, *The waves and the sigmas (to say nothing of the 750 GeV mirage)*, arXiv:1609.01668, <https://arxiv.org/abs/1609.01668>.
- [5] ENFSI, European Network of Forensic Science Institutes, *Guideline for Evaluative Reporting in Forensic Science*, 2016, <http://enfsi.eu/?s=Guideline+for+Evaluative+Reporting+in+Forensic+Science>
- [6] A. Biedermann, C. Champod, T. Hicks and F. Taroni, *Challenging Forensic Science: How Science Should Speak to Court*, <https://www.coursera.org/learn/challenging-forensic-science>

---

<sup>16</sup>For example he derived the formula of the weighted average, stressing its importance to use it, instead of the individual values [12].

<sup>17</sup>Historical French and German translations are also available [15, 16].

<sup>18</sup>Although the formal theory of probability has only been developed in the last few centuries, the noun *probabilitas*, the adjective *probabilis* and especially its comparative *probabilior* (‘more probable’), playing a fundamental role in probabilistic reasoning, were used in Latin with essentially the same meaning we assign to them in ordinary language. For example, a recent ‘grep’ through the Cicero texts collected in The Latin Library [17] resulted in 105 words containing ‘probabil’. And it is rather popular the Cicero’s quote “*Probability is the very guide of life*” [18], although such a sentence does not appear verbatim in his texts, but it is a digest of his thought (see e.g. *De Natura Deorum*, Liber Primus, nr. 12 [17]).

- [7] C.F. Gauss, *Theoria motus corporum coelestium in sectionibus conicis solem ambientum*, Hamburg 1809, [https://archive.org/details/bub\\_gb\\_ORUOAAAQAAJ](https://archive.org/details/bub_gb_ORUOAAAQAAJ).
- [8] C.F. Gauss, *Theory of the motion of the heavenly bodies moving about the sun in conic sections: a translation of Gauss's "Theoria motus."*, translated by Charles H. Davies, 1857, <https://archive.org/details/motionofheavenly00gausrich>.
- [9] G. D'Agostini, *Bayesian reasoning in data analysis – a critical introduction*, World Scientific 2003.
- [10] P.S. Laplace ("M. de la Place"), *Mémoire sur la probabilité des causes par les événements*, Mémoire de l'Académie royale des Sciences de Paris (Savants étrangers), Tome VI, p. 621, 1774, <https://gallica.bnf.fr/ark:/12148/bpt6k77596b/f32>.
- [11] S.M. Stigler, *Laplace's 1774 memoir on inverse probability*, Statistical Science, Vol. 1, No. 3, 359-378, 1986, <https://projecteuclid.org/euclid.ss/1177013620>.
- [12] G. D'Agostini, *On a curious bias arising when the  $\sqrt{\chi^2/\nu}$  scaling prescription is first applied to a sub-sample of the individual results*, arXiv:2001.07562, <https://arxiv.org/abs/2001.07562>.
- [13] S. B. McGrayne, *The theory that would not die: How Bayes' rule cracked the enigma code, hunted down russian submarines, and emerged triumphant from two centuries of controversy*, Yale University Press 2012. (Video of the presentation of the book at Google available at <https://www.youtube.com/watch?v=8oD6eBkjF9o>.)
- [14] G. D'Agostini, *The Fermi's Bayes theorem*, Bulletin of the International Society of Bayesian Analysis 12 (3) 6-8, 2005, <https://bayesian.org/wp-content/uploads/2016/09/0509.pdf>. arXiv:physics/0509080, <https://arxiv.org/abs/physics/0509080>.
- [15] C.F. Gauss, *Théorie du mouvement des corps célestes parcourant des sections coniques autour du soleil*, traduction suivies de notes par Edmond Dubois, 1864, <https://gallica.bnf.fr/ark:/12148/bpt6k77746b/f3.image.r=.langFR>.
- [16] C.F. Gauss, *Theorie der Bewegung der Himmelskörper welche in Kegelschnitten die Sonne umlaufen*, in Deutsche übertragen von Carl Haase, 1865, <https://gdz.sub.uni-goettingen.de/id/PPN60187157X>.
- [17] The Latin Library, <https://www.thelatinlibrary.com/>.
- [18] <https://www.google.com/search?q=probability+is+the+very+guide+of+live&tbm=isch>.