

Bayesian estimate of the Newtonian constant of gravitation

Volker Dose

*Max-Planck-Institut für Plasmaphysik,
EURATOM Association, Boltzmannstrasse 2,
D-85748 Garching bei München, Germany*

(Dated: October 18, 2006)

Keywords: gravitational constant, robust estimation

Abstract

Bayesian probability theory is employed to derive robust, outlier tolerant methods for the estimation of a quantity and the determination of the uncertainty associated with this estimation given a set of data. The procedure is applied to the estimate of the Newtonian constant of gravitation G yielding

$$G = 6.67414(24) \cdot 10^{-11} m^3 kg^{-1} s^{-2}$$

This value is in good agreement with the recently published value of the 2002 CODATA adjustment but offers a four fold reduced, rigorously calculated uncertainty. The uncertainty reflects - unlike results from the conventional least squares analysis - the quoted uncertainties of the data as well as the data scatter.

I. INTRODUCTION

The adjustment of physical constants and conversion factors is an important task relevant for the whole physics community. It is repeatedly carried out by the international CODATA committee. Among all physical constants the estimate of the Newtonian constant of gravitation is a particularly simple problem since this constant does not correlate with any other fundamental constant. Given a set of data $\{d_i\}$ with associated uncertainties $\{\sigma_i\}$ (hereafter also abbreviated by \vec{d} and $\vec{\sigma}$) the Newtonian constant G is - under certain assumptions - estimated by the weighted arithmetic mean well known to every physicist.

$$\hat{G} = u^2(\hat{G}) \cdot \sum_i d_i / \sigma_i^2 \quad , \quad u^2(\hat{G}) = 1 / \sum_i 1 / \sigma_i^2 \quad . \quad (1)$$

This result has quite a startling property. If the data $\{d_i\}$ is stretched with respect to \hat{G} by an arbitrary factor γ

$$d_i \rightarrow d_i + \gamma(d_i - \hat{G}) \quad , \quad (2)$$

the estimate \hat{G} remains the same (which is acceptable) but the uncertainty $u(\hat{G})$ remains unchanged under such a transformation (which is odd). In other words, the uncertainty $u(\hat{G})$ associated with the estimate \hat{G} (2) depends only on the set of data uncertainties but not on the scatter of the data. This is an important deficiency in practice. It will be shown below that (1) rests on the strong assumption that $\{d_i\}$ are samples from Gaussian distributions with mean \hat{G} and exactly known variances σ_i^2 . This assumption is justified only in very rare cases. It will be relaxed below. We shall assume instead that the quoted uncertainties of the data may be wrong by an unknown common factor. Mohr and Taylor [1] have also considered this assumption in an attempt to make the available data consistent. Conceptually this is a highly debatable assumption, but technically it removes the discrepancy between the scatter of the data and the quoted uncertainties. It is the Bayesian treatment of section II which allows to view the estimate (1) as the expectation $E(G)$ of a probability distribution (associated with the given information) and the squared uncertainty $u^2(\hat{G})$ as the variance $var(G)$ of that distribution.

We shall then introduce a generalization of the common distributional assumption. In the sense of the maximum entropy principle [2] the Gaussian form of a distribution results from the knowledge of only mean and variance. An infinite manifold of distributions can be constructed from the knowledge of mean and variance (or standard derivation). A particular

interesting choice is the Laplace distribution. It has of course lower entropy than the Gaussian but offers interesting properties. While in the Gaussian case data point d_i contributes $(d_i - G)^2 / \sigma_i^2$ to the overall misfit, for the Laplace distribution the contribution is $|d_i - G| / \sigma_i$. It is immediately clear that data with values far off the mainstream, which are usually called outliers, have a much more pronounced influence on the misfit for a Gaussian rather than a Laplace likelihood. This argument is of course in favor for the Laplace distribution for all sets of data which do not obey the strong assumption that their sampling distribution is Gaussian.

It would be an uncomfortable situation if we were left with a choice between the two on the basis of a qualitative argument or even as a matter of taste. Fortunately the relative performance can be quantified within the frame work of Bayesian probability theory. The key word is here Bayesian model comparison. We shall also introduce a third likelihood model function into the game which keeps the Gaussian character for the mainstream data and turns gradually over to a Laplace behaviour for more distant measurements. The function offering this is a hyperbolic sech of argument $|d_i - G| / \sigma_i$. Not surprisingly these three different forms of the likelihood function result in general in different pairs of estimates for $E(G)$ and $var(G)$. Within Bayesian probability theory the seeming dilemma of choosing one result out of three is resolved by a weighted mean of the three results with weights equal to the respective model probabilities.

This paper assumes a basic familiarity of the reader with Bayesian probability theory. It is the central tool employed in this paper and a considerable amount of literature on the subject is available. As a good introduction I suggest the text books by Sivia [3] and Gregory [4]. Shorter summaries can be found in the introductory parts of [5] and [6].

II. THE GAUSSIAN LIKELIHOOD

In this section we investigate and extend the Gaussian likelihood. We start with a derivation of (1) employing the Bayesian language. Given a set of data $\{d_i\}$ with exactly known associated variances $\{\sigma_i^2\}$ the least informative (maximum entropy-) distribution encoding this knowledge is a Gaussian [2]

$$p(\vec{d}|G, \vec{\sigma}, I) = \left\{ \prod_i \frac{1}{\sigma_i \sqrt{2\pi}} \right\} \exp \left\{ -\frac{1}{2} \sum_i \left(\frac{d_i - G}{\sigma_i} \right)^2 \right\} . \quad (3)$$

$p(\vec{d}|G, \vec{\sigma}, I)$ is the sampling distribution of the data given G and $\vec{\sigma}$. If considered as a function of G given \vec{d} and $\vec{\sigma}$ it is called the likelihood of G . The likelihood of G is the basic information required to calculate the target quantity $p(G|\vec{d}, \vec{\sigma}, I)$, the probability distribution of G given \vec{d} and $\vec{\sigma}$. The relation between the two distributions is provided by Bayes' theorem.

$$p(G|\vec{d}, \vec{\sigma}, I) = p(G|I) \cdot p(\vec{d}|G, \vec{\sigma}, I) / p(\vec{d}|\vec{\sigma}, I) . \quad (4)$$

$p(G|I)$ is called the prior distribution for G which summarizes the knowledge about G before considering the data. We shall take it flat throughout this paper

$$p(G|I) = \frac{1}{G_{max} - G_{min}} , \quad G_{min} \leq G \leq G_{max} , \quad (5)$$

and zero otherwise. The interval $[G_{min}, G_{max}]$ is assumed to include the true value of G which is unknown. $p(\vec{d}|\vec{\sigma}, I)$ is called the evidence of the data. It follows from the Bayesian marginalization rule

$$p(\vec{d}|\vec{\sigma}, I) = \int p(G|I) \cdot p(\vec{d}|G, \vec{\sigma}, I) dG . \quad (6)$$

The numerical value of $p(\vec{d}|\vec{\sigma}, I)$ is the probability that the data represent fluctuations about an unknown target value G regardless of what the numerical value of G is.

A more convenient form of (3) can be derived using the definitions

$$\frac{N}{\rho^2} = \sum_i \frac{1}{\sigma_i^2} , \quad \frac{N\bar{D}}{\rho^2} = \sum_i \frac{d_i}{\sigma_i^2} , \quad \frac{N\overline{D^2}}{\rho^2} = \sum_i \frac{d_i^2}{\sigma_i^2} . \quad (7)$$

This yields

$$p(\vec{d}|G, \vec{\sigma}, I) = \left(\prod_i \frac{1}{\sigma_i \sqrt{2\pi}} \right) \exp \left\{ -\frac{N}{2\rho^2} (G - \bar{D})^2 - \frac{N}{2\rho^2} \overline{\Delta D^2} \right\} . \quad (8)$$

N is the number of data and $\overline{\Delta D^2}$ the variance defined in the usual way as

$$var(D) = \overline{\Delta D^2} = \overline{D^2} - \bar{D}^2 . \quad (9)$$

The posterior distribution of G given \vec{d} and $\vec{\sigma}$ is then

$$p(G|\vec{d}, \vec{\sigma}, I) = \frac{p(G|I)}{p(\vec{d}|\vec{\sigma}, I)} \cdot \left(\prod_i \frac{1}{\sigma_i \sqrt{2\pi}} \right) \exp \left\{ -\frac{N}{2\rho^2} (G - \bar{D})^2 - \frac{N}{2\rho^2} \overline{\Delta D^2} \right\} . \quad (10)$$

The posterior distribution of G given \vec{d} and $\vec{\sigma}$ is the most detailed result of a Bayesian analysis. It is sometimes both, possible and desirable to characterize it in terms of a few

numbers. We use throughout this paper mean and variance of the posterior distribution. Other choices as for example mode or median and arguments of tail areas are also in use and may be more informative for skew distributions. Mean and variance of (10) are given by

$$E(G) = \overline{D} , \quad var(G) = \rho^2/N , \quad (11)$$

and we are back at the results already cited in the introduction (1). It is interesting to note, that the exponential in (10) depends on the data only through \overline{D} , ρ^2 and $\overline{D^2}$. The set \overline{D} , ρ^2 and $\overline{D^2}$ is called a sufficient statistic of the data in a Gaussian likelihood. A sufficient statistic smaller than the number of data exists only in exceptional cases and the Gaussian form of the likelihood (10) is an example.

The weakness of (11) has already been discussed in the introduction. A possible modification of the likelihood (3, 8) consists of introducing a scaling factor α common to all uncertainties but of unknown magnitude. Such a factor, if properly chosen reduces the inconsistency of the quoted data uncertainties and the data scatter. The modified likelihood, which we mark by an additional discrete model variable M_1 is then

$$p(\vec{d}|G, \vec{\sigma}, \alpha, M_1, I) = \left(\prod_i \frac{1}{\sigma_i \sqrt{2\pi}} \right) \frac{1}{\alpha^N} \exp \left\{ -\frac{N}{2\alpha^2 \rho^2} (G - \overline{D})^2 - \frac{N}{2\alpha^2 \rho^2} \overline{\Delta D^2} \right\} . \quad (12)$$

The joint posterior distribution for G and α becomes using Bayes' theorem

$$p(G, \alpha | \vec{d}, \vec{\sigma}, M_1, I) = p(\alpha | I) \cdot p(G | I) \cdot p(\vec{d} | G, \vec{\sigma}, \alpha, M_1, I) / p(\vec{d} | \vec{\sigma}, M_1, I) . \quad (13)$$

(13) requires the specification of a further distribution $p(\alpha | I)$. In the absence of any information about the magnitude of α prior to taking notice of the data, we use the transformation invariant uninformative Jeffreys' prior [3]

$$p(\alpha | B, I) = \frac{1}{2 \ln B} \cdot \frac{1}{\alpha} , \quad \frac{1}{B} \leq \alpha \leq B . \quad (14)$$

In a strict sense (14) is only transformation invariant for support parameter $B \rightarrow \infty$. In this limit (14) is no longer normalizable and becomes improper. At this stage we stay with the bounded normalized form (14) and discuss the limit where appropriate. The joint posterior for G and α is now fully specified and can be used to calculate moments

$$E(G^m \alpha^n) = \int \alpha^n d\alpha \int G^m dG p(G, \alpha | \vec{d}, \vec{\sigma}, M_1, I) . \quad (15)$$

Interesting cases are $(m = 0, 1, 2, n = 0)$ und $(m = 0, n = 0, 1, 2)$ to estimate means and variances for G und α . Let us now assume that G_{min}, G_{max} , and B have been chosen so large, that the limits of integration in (15) can be extended $-\infty < G < \infty$ and $0 \leq \alpha < \infty$ without changing the values of the definite integrals in (15) to any prescribed precision. Both integrations can then be done analytically. Imagine to perform the α -integration first. It will destroy the Gaussian structure of the likelihood since the result is a student distribution. It has the same maximum position $G = \overline{D}$ as (12) but a width which is now proportional to $(\overline{\Delta D^2})^{1/2}$ rather than ρ . The step of α -integration couples the data scatter and the data uncertainties. Results of a subsequent G -integration are then

$$E(G) = \overline{D} \quad , \quad var(G) = \frac{\overline{\Delta D^2}}{N - 3} . \quad (16)$$

Analogous estimates for α are

$$E(\alpha) = \left(\frac{N \overline{\Delta D^2}}{2\rho^2} \right)^{1/2} \cdot \Gamma\left(\frac{N-2}{2}\right) / \Gamma\left(\frac{N-1}{2}\right) ,$$

$$var(\alpha) = \frac{N \overline{\Delta D^2}}{2\rho^2} \left\{ \frac{2}{N-3} - \left[\frac{\Gamma(\frac{N-2}{2})}{\Gamma(\frac{N-1}{2})} \right]^2 \right\} . \quad (17)$$

(16) and (17) have been applied to the data set used for the CODATA_98 evaluation [7]. The reason that we stay at this stage with the CODATA_98 data set is that the CODATA_02 data set contains conceptual mistakes which will be discussed and corrected after the Bayesian procedures have been established and tried on the '98 data set. Moreover, the '98 data set contains a measurement far off the mainstream [8] such that it was neglected in the '98 evaluation of the point estimate. It gave, however, rise to an arbitrary enhancement of the calculated uncertainty by the CODATA committee. The aim of the present paper is to make data censoring and ad hoc adjustments superfluous by developing sufficiently robust methods. In order to avoid misunderstandings I want to emphasize at this point that I keep the highest regard for the painstaking, formidable CODATA task of evaluating the validity of experiments performed at various excellent laboratories around the world. My point is different: once an experiment has been identified as valid and state of the art, the numerical result of its measurements, point estimate and uncertainty should be accepted.

The first row in Table I shows the least squares result (1). The second row shows results from (16) and (17). The very large expectation value $E(\alpha)$ shows the extreme inconsistency

TABLE I: Point estimates and uncertainties of the gravitational constant employing the likelihoods proposed in this paper. Data are taken from [7].

MODEL	$\frac{10^{11}G}{m^3kg^{-1}s^{-2}}$	$E(\alpha)$	Mod-prob
LSQF	6.6818(3)	-	-
GAUSS	6.6818(65)	24.7(6.7)	0.002
LAPLACE	6.6739(23)	10.5(3.9)	0.418
COS H ⁻¹	6.6751(41)	11.4(4.1)	0.580

of this data set due to the far away high precision measurement of Michaelis et al. [8]. This is also reflected in the increase of the uncertainty of G by a factor of 22. The last entry in line 2 displays the probability that the model employing [1] explains the data satisfactorily when compared to the models employing the alternative likelihood functions discussed in the following two sections. The calculation of these numbers will be addressed in section V.

III. THE LAPLACE LIKELIHOOD

The information assumed available, mean and variance, can of course be used to specify alternative distributions albeit with lower entropy than the Gaussian. We label with M_2 a distribution which uses the modulus rather than the square as a distance measure. This is, properly normalized

$$p(\vec{d}|G, \vec{\sigma}, \alpha, M_2, I) = \left(\prod_i \frac{1}{2\sigma_i} \right) \frac{1}{\alpha^N} \exp \left\{ -\frac{1}{\alpha} \sum_i \left| \frac{d_i - G}{\sigma_i} \right| \right\}. \quad (18)$$

The posterior for G and α is again obtained employing Bayes' theorem with assumptions on $p(G|I)$ and $p(\alpha|I)$ as before. The calculation of moments from the posterior is not as simple as for M_1 . But surprisingly it can be done analytically. To understand this, we point out that the log-posterior is a polygon as a function of G . Assume now a re-labeling of the data such that $d_i < d_{i+1}$ for all i . The case $d_k = d_{k+1}$ can be absorbed by dropping d_{k+1} and changing $1/\sigma_k$ to $1/\sigma_k + 1/\sigma_{k+1}$. We see then that for $G < d_1$ the log-posterior is a straight

line as a function of G with slope up to a factor of $1/\alpha$

$$s_0 = \sum_i 1/\sigma_i . \quad (19)$$

For the interval $d_1 \leq G \leq d_2$ data point d_1 contributes a negative slope $-1/\sigma_1$ and all the other data positive values $1/\sigma_i$. We have in general

$$s_1 = s_0 - 2/\sigma_1 , \quad s_i = s_{i-1} - 2/\sigma_i , \quad i = 2 \dots N . \quad (20)$$

Alternatively the slope of the polygon in G in the interval $[d_i, d_{i+1}]$ can be expressed as

$$s_i = \frac{f_{i+1} - f_i}{d_{i+1} - d_i} , \quad f_i = - \sum_{k \neq i} \left| \frac{d_i - d_k}{\sigma_k} \right| . \quad (21)$$

The likelihood for $d_i \leq G \leq d_{i+1}$ can therefore be written as

$$p(\vec{d}|G, \vec{\sigma}, \alpha, M_2, I) = \frac{Z_2}{\alpha^N} \exp \left\{ \frac{f_i}{\alpha} + \frac{s_i}{\alpha} (G - d_i) \right\} , \quad (22)$$

with Z_2 given by

$$Z_2 = \left(\prod_i \frac{1}{2\sigma_i} \right) . \quad (23)$$

G -integrals over the posterior reduce then to the type

$$I_i(\alpha, k) = \frac{Z_2}{\alpha^N} \exp \left\{ \frac{f_i}{\alpha} - \frac{s_i d_i}{\alpha} \right\} \int_{d_i}^{d_{i+1}} G^k dG \exp \left\{ \frac{s_i}{\alpha} G \right\} , \quad (24)$$

which is again elementary. The alternative expression for s_i in (21) is useful to simplify the result of this integration, yielding for the simplest case $k = 0$

$$I_i(\alpha, 0) = \frac{Z_2}{\alpha^{N-1}} \frac{1}{s_i} \left\{ \exp \left\{ \frac{f_{i+1}}{\alpha} \right\} - \exp \left\{ \frac{f_i}{\alpha} \right\} \right\} . \quad (25)$$

The remaining α -integral can be reduced to a Γ -function keeping in mind that all $f_i < 0$ and yields

$$I_i(0) = Z_2 \cdot \Gamma(N - 1) \cdot \left\{ |f_{i+1}|^{-(N-1)} - |f_i|^{-(N-1)} \right\} . \quad (26)$$

For the end intervals $G \leq d_1$ and $G \geq d_N$ one of the limits of integration in (24) becomes infinite. The results for these cases are

$$I_0(0) = \frac{Z_2}{s_0} \cdot \frac{\Gamma(N - 1)}{|f_1|^{N-1}} , \quad I_N(0) = \frac{Z_2}{|s_N|} \frac{\Gamma(N - 1)}{|f_N|^{N-1}} . \quad (27)$$

Higher moments for G and α are of course more cumbersome but can be evaluated along the indicated scheme.

Results for the estimate of G using the '98 data set and model M_2 , the Laplace likelihood, are shown in the third row of Table I. The most remarkable result is the drastic reduction of the mean value of G , but note that also the uncertainty has decreased by a factor of 3. The estimate of G is in fact compatible with the result obtained by the CODATA group neglecting the problematic measurement [8] and is convincing demonstration of the robustness of (18) against outliers. We see further, that the expectation value of α has decreased by a factor of 2.3 and the probability of this model has strongly increased compared to M_1 .

IV. THE HYPERBOLIC COSINE LIKELIHOOD

Interestingly there exists a likelihood function which exhibits the small argument character of a Gaussian and the large argument character of the Laplace density. Consider the normalized function

$$f(x|\alpha) = \frac{2}{\pi\alpha} \cdot \frac{1}{\exp(|x|/\alpha) + \exp(-|x|/\alpha)}, \quad x = \left| \frac{d - G}{\sigma} \right|. \quad (28)$$

Expanding the denominator we obtain

$$f(x|\alpha) \approx \frac{1}{\pi\alpha} \cdot \frac{1}{1 + \frac{1}{2} \left| \frac{x}{\alpha} \right|^2} \approx \frac{1}{\pi\alpha} \left(1 - \frac{1}{2} \left| \frac{x}{\alpha} \right|^2 \right). \quad (29)$$

This must be compared to the small argument expansion of a normalized Gaussian

$$\frac{1}{\alpha\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left| \frac{x}{\alpha} \right|^2 \right\} = \frac{1}{\alpha\sqrt{2\pi}} \left(1 - \frac{1}{2} \left(\frac{x}{\alpha} \right)^2 + \dots \right). \quad (30)$$

For N independent data the likelihood based on (30) becomes

$$p(\vec{d}|G, \vec{\sigma}, \alpha, M_3, I) = \left(\prod_i \frac{1}{2\sigma_i} \right) \frac{1}{\alpha^N} \exp \left\{ -\frac{1}{\alpha} \sum_i \left| \frac{d_i - G}{\sigma_i} \right| \right\} \cdot \left(\frac{4}{\pi} \right)^N \prod_k \left\{ 1 + \exp \left[-\frac{2}{\alpha} \left| \frac{d_k - G}{\sigma_k} \right| \right] \right\}^{-1}. \quad (31)$$

(31) has been arranged such that the first factor becomes identical to the likelihood M_2 and the second factor ψ provides a correction which is bounded

$$\left(\frac{2}{\pi} \right)^N \leq \psi \leq \left(\frac{4}{\pi} \right)^N. \quad (32)$$

I was unable to find a way to obtain analytic integrals of the posterior constructed from (31) and its moments in α and G . These computations had therefore to be done numerically. The availability of analytic results for models M_1 and M_2 was of great help in the course of this work because the accuracy of the numerical procedure could be calibrated by comparison to the respective analytic results. Application of model M_3 to the CODATA_98 data set yielded the results shown in the fourth row of Table I. The mean has slightly increased such that it lies between Gauss and Laplace results. The same holds for the uncertainty. The expectation value $E(\alpha)$ shows also a ten percent increase. All these numbers are in line with the property of M_3 that it approaches Gaussian behavior for small arguments and turns over to Laplace behavior for large arguments. The model probability for M_3 is highest among the three signalling that the likelihood M_3 is best suited to represent the data.

V. AVERAGE OVER MODELS

At this stage we face the problem that we have three different answers of the estimation problem related to the three different likelihoods. The discrete variable M_i can however be eliminated employing the Bayesian marginalization rule. The posterior density which we should have used in the first place is $p(\alpha, G|\vec{d}, \vec{\sigma}, I)$. It can be obtained from the identity

$$p(\alpha, G|\vec{d}, \vec{\sigma}, I) = \sum_i p(\alpha, G, M_i|\vec{d}, \vec{\sigma}, I) , \quad (33)$$

which decomposes upon use of the product rule into

$$p(\alpha, G|\vec{d}, \vec{\sigma}, I) = \sum_i p(M_i|\vec{d}, \vec{\sigma}, I) \cdot p(\alpha, G|\vec{d}, \vec{\sigma}, M_i, I) . \quad (34)$$

The expansion coefficients $p(M_i|\vec{d}, \vec{\sigma}, I)$ are related to the marginal likelihoods $p(\vec{d}|\vec{\sigma}, M_i, I)$ via Bayes' theorem

$$p(M_i|\vec{d}, \vec{\sigma}, I) = p(M_i|I)p(\vec{d}|\vec{\sigma}, M_i, I)/p(\vec{d}|\vec{\sigma}, I) . \quad (35)$$

Use of (35) requires to specify prior probabilities $p(M_i|I)$. The most uninformative choice is uniform $p(M_i|I) = 1/3$ for all i . The expansion coefficients can then be obtained from previously calculated evidences. Numerical values of the model probabilities are given in the last column of Table I and Table IV. The final result for the expectation value of $\alpha^n G^m$

TABLE II: Point estimates and uncertainties from the Bayesian approach compared to the least squares estimate.

Source	$\frac{10^{11}G}{m^3kg^{-1}s^{-2}}$	Key
CODATA_98 adjustment	6.6729(100)	(a)
This work, same data set	6.6746(35)	(b)
This work deleting [8]	6.6732(7)	(c)

is then

$$E(\alpha^n G^m) = \sum_i p(M_i | \vec{d}, \vec{\sigma}, I) \cdot E(\alpha^n G^m | \vec{d}, \vec{\sigma}, M_i). \quad (36)$$

Note that we must first use (36) to find moments from which we subsequently calculate variances and uncertainties. Table II shows final results for the CODATA_98 data set. The first and third entries are essentially identical apart from the uncertainty which was arbitrarily estimated in the CODATA_98 adjustment while it was calculated in this work. Most interesting is the middle entry which includes the outlier [8] in the analysis with little influence on the point estimate. A strong effect is seen in the uncertainty, which rises by a factor of five, reflecting the fact that the [8] is indeed highly inconsistent with the rest of the data. Quite the same, this result is an impressive demonstration of the robustness of the method developed in this paper.

A. The CODATA_02 adjustment

The CODATA_02 evaluation rests on ten different data entries. Two of them are the same as in the '98 data set. One new data point stems from Gundlach and Merkowitz [9]. Five data are from experiments which have been improved since the '98 adjustment. Finally the remaining two "data" were chosen as the 1986 and 1998 CODATA recommended values. This is in order as far as the 1986 recommended value [10] is concerned since this was a single measurement whose uncertainty was, however, arbitrarily doubled by the committee. The inclusion of the '98 estimate, however, cannot be accepted for two reasons. The first is that this kind of information transfer holds only in the case of a Gaussian likelihood with (assumed) exact uncertainties of the data. It fails already for our likelihood M_1 let

alone M_2 and M_3 . The second reason is of conceptual nature. The '98 evaluation rests on ten data, out of which seven appear in the '02 evaluation, two unchanged and five from improved experiments. Inclusion of the information of the two unchanged data also via the '98 adjusted value puts an unjustified higher weight on them. Inclusion of the five data from different stages of the same experiment, the '98 data via the '98 adjustment and the '02 data as the new data violate in addition the independence assumption which is the basis for the product form of the likelihoods even for the simplest case (3). This work employs therefore a modified data set for the final evaluation of G . The 1986 data point was used with its originally published uncertainty [11]. The data from the JILA_98 experiment [12] was included. The CODATA committee chose to exclude it from the '02 adjustment "because of its relatively large uncertainty" [1]. This argument appears rather odd to me. It probably dates back to the age of mechanical desk calculators. Fortunately modern computing facilities do not limit the size of a data set. Moreover in the light of the previously developed likelihood models both pieces of information, deviation from the mainstream and quoted uncertainty are important for robust estimation of G . We do not include the measurement [8] which was discordant with the rest of the '98 data set because a review of various critical aspects of the experiment led Michaelis et al. to withdraw their former result [13]. In summary, the data set underlying the CODATA_02 evaluation was modified by withdraws of the '98 recommended value, addition of the JILA_98 [12] data and restoration of the uncertainty of the CODATA_86 recommended value. In order to avoid ambiguities we reproduce the modified data set which forms the basis for the final result of this paper in Table III.

Table IV shows results for the different models discussed in this paper. The point estimates agree all very well with each other. The difference appears in the uncertainty. Clearly the least squares result overestimates the precision. Remember that this algorithm is insensitive to the scatter of the data. The expectation values $E(\alpha)$ show that this scatter is in fact larger than tacitly assumed ($\alpha = 1!$) in the least squares algorithm. However, altogether this data set is of much higher quality than the '98 data and of similar quality (result not shown) if the data [8] is removed from the '98 data set. The model probabilities for M_1 and M_2 are quite similar indicating that both aspects are contained in the data set. It is therefore most satisfactory that model M_3 which addresses both aspects turns out to be the most probable. Table V shows final results after marginalization over the models. The three estimates are

TABLE III: Data set employed for the final Bayesian estimate of G . This is a modification of the data set used in [1].

Identification	Reference	$\frac{10^4 G}{m^3 kg^{-1} s^{-2}}$
Gundlach and Merkowitz	[9]	6.674255(92)
Luther and Towler	[11]	6.67259(43)
Schwarz et al.	[12]	6.6873(94)
Bagley and Luther	[14]	6.67400(70)
Karagioz	[15]	6.67290(50)
Luo et al.	[16]	6.67090(70)
Quinn et al.	[17]	6.67559(27)
Kleinevoß	[18]	6.67422(98)
Schlamminger et al.	[19]	6.67407(22)
Armstrong and Fitzgerald	[20]	6.67387(27)

TABLE IV: Performance of the likelihoods developed in this paper for the latest data set given in Table III.

MODEL	$\frac{10^{11} G}{m^3 kg^{-1} s^{-2}}$	$E(\alpha)$	Mod-prob
LSQF	6.67419(7)	-	-
GAUSS	6.67419(24)	3.1(8)	0.22
LAPLACE	6.67411(20)	2.4(9)	0.29
COS H^{-1}	6.67414(25)	2.3(8)	0.49

TABLE V: Point estimates and uncertainties from the full Bayesian calculation compared to the CODATA_02 adjustment [1].

Source	$\frac{10^{11} G}{m^3 kg^{-1} s^{-2}}$	Key
CODATA_02 adjustment	6.67422(100)	(d)
This work, same data set	6.67416(21)	(e)
This work modified set	6.67414(24)	(f)

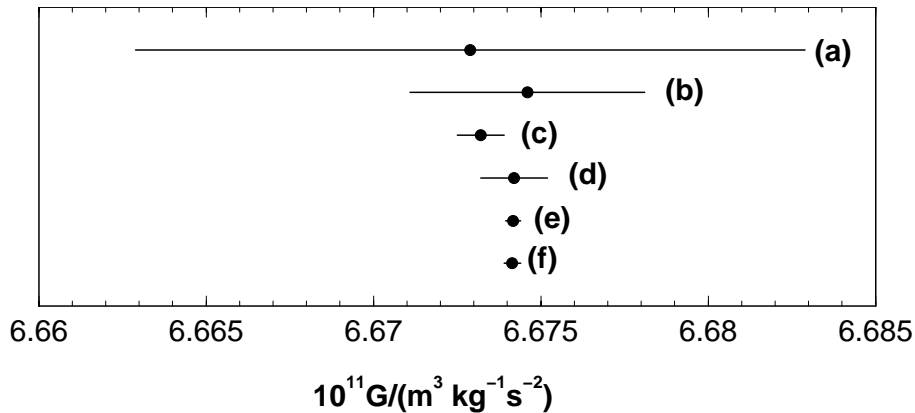


FIG. 1: Results of the present work (b, c, e, f) in comparison to the CODATA_98 (a) and CODATA_02 (d) recommendation. Labels a-f correspond to the keys given in Table II and Table V.

highly consistent and the main improvement over the CODATA_02 adjustment appears to be the narrowing of the uncertainty by a factor of four. The recommended value of G on the basis of the algorithm developed above is

$$G = 6.67414(24) \cdot 10^{-11} m^3 kg^{-1} s^{-2} . \quad (37)$$

It may please the proponents of the least squares algorithm as the ultimate data analysis tool to realize that the result (37) coincides nearly with the CODATA_02 least squares estimate. The coincidence is, however, accidental since the two evaluations differ in data sets and methods. Remember also, that (37) is characteristic for the given data set while identical least squares results can be obtained from an infinite manifold of data sets featuring the same sufficient statistic ρ, \vec{D} . We recall further that (37) provides a rigorous uncertainty estimate which relies on both, the scatter of the data and the quotation of their experimental uncertainties. All results collected in Tables II and V are also shown in Fig. 1 in order to facilitate comparison.

VI. CONCLUSION

The central claim of this paper is that the method developed is very robust compared to the physicists pet tool, the least squares analysis. The robustness of the presented methods is further corroborated by a data simulation. The upper panel of Fig. 2 shows some of the

data on which the CODATA_98 was based as full dots with associated uncertainties. The open circle represents an additional fictitious measurement which is moved across the cloud of full dots in the range indicated by the small dots. The lower panel shows the resulting estimate of G (vertical scale) obtained from the Gaussian(dash-dotted)-, the Laplace(dotted)- and the \cosh^{-1} (dashed)-likelihoods as well as the model averaged total as the continuous trace. While estimates based on the Gaussian likelihood exhibit a linear response of the estimate on the position of the fictitious measurement, the other models show a clear saturation when the fictitious measurement becomes more of an outlier. This is conclusive proof of their robustness. Given this robustness, data censoring preceding the analysis is no longer necessary nor desirable. Of course, a censoring of the whole experiment and procedure leading to a particular numerical value of the quantity in question is absolutely necessary. This is, however, something entirely different than a census based on the amount of agreement of a particular outcome with the rest of the data under consideration. The CODATA committee has had anyway bad experience with excluding data on the grounds that they seem to be discrepant with the rest. A prominent example is the difference between the 1973 and 1986 adjustments of the Faraday constant. Much of this discrepancy is due to exclusion of two seemingly discrepant measurements in the 1973 evaluation [21]. The lesson for the Newtonian constant of gravitation is that there was insufficient reason in the 1998 adjustment to exclude the PTB_95 data. Conversely, there was very good reason to exclude it from the 2002 adjustment because it had meanwhile turned out that several aspects of the experiment appeared doubtful and were no longer supported by the authors. The second aspect of robustness concerns uncertainties. There are numerous examples in the CODATA reports that uncertainties were ad hoc enlarged because the result of the least squares procedure was considered improper in view of the observed data scatter. This holds also for the '98 and '02 adjustments of the Newtonian constant of gravitation, see first rows of Table II and Table IV. The methods given in this paper consider in an integrated way throughout simultaneously the quoted uncertainties and the scatter of the data. A beautiful example of how this works is given in Table II. The third row shows the estimate from the present work when the PTB_95 data is deleted from the set as was done in the CODATA adjustment. The second row shows the result without an - at that time - unjustified data censoring. While the point estimate shows a marginal change only the uncertainty increases by a factor of five.

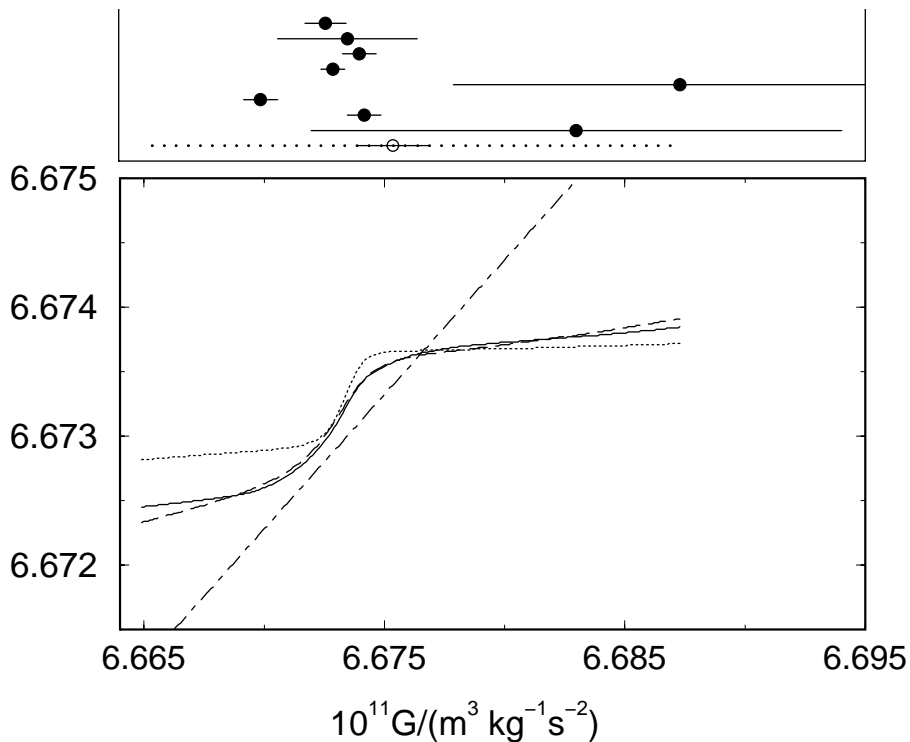


FIG. 2: Performance of the different likelihoods employed in this work when a fictitious measurement shown as an open circle is moved across a set of data shown as full dots. Dash dotted: Gaussian-likelihood, dotted: Laplace-likelihood, dashed: \cosh^{-1} -likelihood. The continuous trace shows the model averaged estimate. The vertical scale refers to the estimated value of G as a function of the position of the fictitious measurement.

The present case of data analysis is a particularly simple one, since G does not correlate with any other physical constant. It is therefore very well suited to develop methods because the interpretation of results is simplest in this case. However, there is not the slightest reason why application could not be extended to the multivariate problem. The basic deficiency of the least squares method persists in dimensions greater than one but becomes much less obvious.

A good summary of the above arguments is the following citation from the New Scientist [22]: "We use fantastic telescopes, the best physical models and the best computers. The weak link in this chain is interpreting our data using 100-year-old mathematics." This must change, the sooner, the better.

I conclude with a final comment on the assumption used throughout this paper that

the uncertainties of the various measurements are in error by an unknown common factor. The main reason for this was the corresponding attempt by Mohr and Taylor in order to achieve consistency of the data set. While this assumption sounds reasonable for a bunch of measurements from a single laboratory, it is in fact rather unrealistic for a set of data from various different laboratories. A much more reasonable assumption for this situation is that quoted uncertainties should not be taken to be the true uncertainties but rather point estimates of the true uncertainties. A re-evaluation of the revised data set for G in Table III based on this modified assumption concerning quoted uncertainties is currently performed.

- [1] P. J. Mohr and B. N. Taylor. *Rev. Mod. Phys.*, 77:1, 2005.
- [2] J. N. Kapur and H. K. Kesavan. *Entropy Optimization Principles with Applications*. Academic, Boston, 1992.
- [3] D. S. Sivia. *Data Analysis: A Bayesian Tutorial*. Clarendon Press, Oxford, 1996.
- [4] P. C. Gregory. *Bayesian Logical Data Analysis for the Physical Sciences; A Comparative Approach with Mathematica Support*. Cambridge University Press, Boston, 2005.
- [5] P. C. Gregory and T. J. Loredo. *Astrophys. J.*, 398:142, 1992.
- [6] V. Dose. *Rep. Prog. Phys.*, 66:1421, 2003.
- [7] P. J. Mohr and B. N. Taylor. *Rev. Mod. Phys.*, 72:351, 2000.
- [8] W. Michaelis, H. Haars, and R. Augustin. *Metrologia*, 32:267, 1996.
- [9] J. H. Gundlach and S. M. Merkowitz. *Phys. Rev. Lett.*, 85:2869, 2000.
- [10] E. R. Cohen and B. N. Taylor. *Rev. Mod. Phys.*, 59:1121, 1987.
- [11] G. G. Luther and W. R. Towler. *Phys. Rev. Lett.*, 48:121, 1982.
- [12] J. P. Schwarz, D. S. Robertson, T. M. Niebauer, and J. E. Faller. *Science*, 282:2230, 1998.
- [13] W. Michaelis, J. Melcher, and H. Haars. *Metrologia*, 41:L29, 2004.
- [14] C. H. Bagley and G. G. Luther. *Phys. Rev. Lett.*, 78:3047, 1997.
- [15] O. V. Karagioz, V. P. Izmaylov, and G. T. Gillies. *Grav. Cosmol.*, 4:239, 1998.
- [16] J. Luo, Z. K. Hu, X. H. Fu, S. H. Fan, and M. X Tang. *Phys. Rev. D*, 59:042001, 1999.
- [17] T. J. Quinn, C. C. Speake, S. J. Richman, R. S. Davis, and A. Picard. *Phys. Rev. Lett.*, 87:111101, 2001.
- [18] U. Kleinevoß. Ph. D. Thesis (2002), Bestimmung der Newtonschen Gravitationskonstanten G, University of Wuppertal.
- [19] S. Schlamminger, E. Holzschuh, and W. Kündig. *Phys. Rev. Lett.*, 89:161102, 2002.
- [20] T. R. Armstrong and M. P. Fitzgerald. *Phys. Rev. Lett.*, 91:201101, 2003.
- [21] E. R. Cohen and B. N. Taylor. *J. Phys. Chem. Ref. Data*, 2:663, 1973.
- [22] New Scientist, 26 June 2004.