Measurement 94 (2016) 26–30

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Model uncertainty and reference value of the Planck constant

G. Mana

INRIM – Istituto Nazionale di Ricerca Metrologica, str. delle Cacce 91, 10135 Torino, Italy

ARTICLE INFO

Article history:
Received 3 May 2015
Accepted 18 July 2016
Available online 20 July 2016

MSC:
62F15
62F03

Keywords:
Bayesian inference
Hypothesis testing
Determination of fundamental constants
Probability theory

ABSTRACT

Statistical parametric models are proposed to explain the values of the Planck constant obtained by comparing electrical and mechanical powers and by counting atoms in ^{28}Si enriched crystals. They assume that uncertainty contributions – having heterogeneous, datum-specific, variances – might not be included in the error budgets of some of the measured values. Model selection and model averaging are used to investigate data consistency, to identify a reference value of the Planck constant, and to include the model uncertainty in the error budget.

1. Introduction

The definition of a system of units on the basis of conventional values of fundamental constant of physics [1] is motivating efforts on determinations of the Planck constant [2]. The most accurate data come from the comparison of mechanical and electrical powers by watt-balance experiments [3,4] and the count of the atoms in ^{28}Si enriched silicon balls [5]. Four h determinations comply with the accuracy required to make the kilogram redefinition feasible [6–10]. A statistical analysis of these results is necessary to check their consistency and to chose a reference value of the Planck constant.

Data analysis is usually carried out by selecting a model and by processing the measurement results as if they had generated by it. This approach ignores the model uncertainty and can lead to underestimates of the uncertainty, to overconfident inferences, and to decisions that are more risky than one thinks they are. Questions are: How accurately does a model explain the data and what is the impact of the model uncertainty on the measurand estimate and the inferences that we draw from the measurement results? Given an uncertain data model and a measurand estimate based on it, how can the total uncertainty of the measurand value be assessed?

Probability calculus can select the model most supported by the data and include the uncertainty into the analysis and uncertainty budget [11]; an example investigating the choice of the degree when fitting a polynomial to noisy data is given in [12]. The choice of a measurand value from inconsistent data-sets is investigated in [13–16].

This paper builds on these works and delivers some additional results. Firstly, it considers models where the standard deviations of a data subset – the empty set and the whole data set included – might be larger than the associated uncertainties; but, we do not know what this subset is. Secondly, it chooses the uninformative prior distribution of the unknown standard-deviations by requiring that Gaussian sampling-distributions of the measurement results are equiprobable. A novelty is that, if these standard-deviations are not of interest, marginalization allows an analytical expression of the measurement-result distributions to be given, no matter what the standards deviations – greater than or equal to the associated uncertainties – may be. Eventually, since one of the subset does apply, this paper tests the data consistency by comparing the probability of each subset is the right one given the data and suggests a reference value of the Planck constant by averaging over all the subsets. In this way, all the data determine the reference value, no measurement result is excluded, and none is considered fully reliable or suspicious.

2. Planck constant values

The starting point of the analysis is the list in Table 1. In 2014, the Bureau International des Poids et Mesures (BIPM) carried out a

E-mail address: g.mana@inrim.it

http://dx.doi.org/10.1016/j.measurement.2016.07.056
0263-2241/

Table 1
Measured values of the Planck constant; f is the fraction of the systematic contribution to the uncertainty budget.

Lab	Year	Reference	Label	$10^{34}h$ (J s)	f
IAC[a]	2011	[6,7]	–	6.62606991(20)	
IAC[a]	2015	[7]	–	6.62607016(13)	
IAC[b]	2015	[19]	1	6.62607009(12)	0.16
NIST	2015	[8]	2	6.62606936(37)	0.05
NRC	2014	[9,10,7]	3	6.62607011(12)	0.17
		This paper	–	6.626070073(94)	

IAC – International Avogadro Coordination.
NIST – National Institute of Standards and Technology (USA).
NRC – National Research Council (Canada).
[a] These values' correlation is 17% [19].
[b] Average of the 2011 and 2015 IAC's correlated-values.

campaign of mass calibration with respect to the international prototype, in anticipation of the redefinition of the kilogram [17]. This brought to light an offset of the BIPM as-maintained mass unit, which was traceable to the prototype in 1992. Therefore, the mass values used in the watt-balance and atom counting experiments, were suitably corrected.

The IAC's N_A values are converted into Planck constant values via the molar Planck constant $N_A h = 3.9903127176(28) \times 10^{-10}$ J s mol^{-1}, which has a negligible uncertainty [18]. The correlation of the N_A values reported in 2011 and 2015 by the IAC is investigated in [19], which gives also the mean of the correlated values. To avoid complications due to the correlation, the input datum for this analysis is the mean of the 2011 and 2015 IAC's values.

The values selected for this analysis are labelled from 1 to 3 in Table 1; they are shown in Fig. 1. The BIPM estimated the calibration uncertainty as 3 µg; this uncertainty affects all the mass values in the watt-balance and atom counting experiments. The Table 1 gives the fractions of this systematic component of the uncertainty budget; the correlation of any pair of h values can be obtained by multiplying the pair's systematic fractions.

In 2012, the consultative committee for mass and related quantities of the International Committee for Weights and Measures recommended that "*...the values provided by the different experiments be consistent at the 95% level of confidence*" [20]. Since the confidence level is a concept associated to the Neyman's confidence interval [21,22], the meaning of this recommendation is not very clear.

A way to examine the data consistency might be the significance test of Fisher [11]. Assuming that the data are independent normal-variables having the same mean h and standard deviations

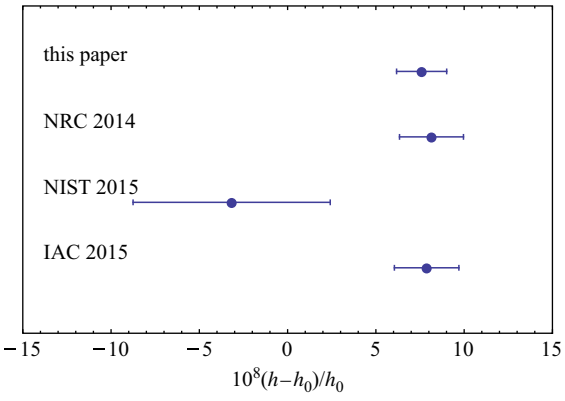

Fig. 1. Measured values of the Planck constant; $h_0 = 6.62606957 \times 10^{-34}$ J s is the value recommended by the Committee on Data for Science and Technology.

σ_i equal to the associated uncertainties, u_i – which is the consistent-data or null hypothesis, a test statistic is the Pearson χ^2 variable [23]. By choosing a 5% significance level, the expected 95% quantile is $\chi^2_{0.95} = 6.0$. For this data set, the observed value is $\chi^2 = 3.8$; since this value is less than the $\chi^2_{0.95}$ rejection level, the consistent-data hypothesis is accepted. The test ensures that the probability of rejecting the consistent-data model when it is true is 5%, but to accept the consistent-data hypothesis as correct is an *argumentum ad ignorantiam* fallacy.

In order to assess the data consistency, we must calculate the probability of the null hypothesis; this requires to include it into a wider hypothesis space. To this end, we consider underestimations of the data uncertainties. Accordingly, each datum x_i is thought to be a random variable having mean h and variance $\sigma_i^2 = u_i^2 + \lambda_i^2$, where, when reporting the measurement uncertainty, a datum-specific contribution to the variance, λ_i^2, was omitted. It is also possible that some measurement uncertainty was correctly evaluated – that is, for these measurements, $\sigma_i = u_i$. Of course, all the measurement uncertainty might be correctly evaluated.

Our assumption is that there exists a subset of good data – which might be the empty set or the full data list – having correct uncertainty assessments; its elements x_i are independent realizations of random variables having variances u_i^2. For the remaining data, the uncertainties u_i are only lower bounds to the standard deviations, that are additional model parameters. The good data can not be predetermined; instead, all the subsets will be taken in turn as the sought good-data subset. The final h estimate will be obtained by model averaging using the probability of the each subset being the good one.

3. Theoretical framework

Before going into the specific application to the data in Table 1, this section outlines the theoretical framework of the analysis.

3.1. Model selection

In order to explain the measurement results, we consider a number of parametric statistical models – say, $M_A, M_B, \ldots$ – where each model is parameterized by the measurand h and, possibly, a set of nuisance parameters $\boldsymbol{\sigma}$. We assume that the models are mutually exclusive and complete, that is, $\bigvee_n M_n = \text{True}$. Given the measurement results $\boldsymbol{x} = [x_1, x_2, x_3]^{\mathrm{T}}$ and the data likelihood $L(\boldsymbol{x}|h, \boldsymbol{\sigma}, M)$, one proceeds by assigning a prior probability distribution $\pi(h, \boldsymbol{\sigma}|M)$ to the model parameters and a prior probability $\Pi(M)$ to each model. Next, by using the product rule of the probabilities, the joint distribution of the data, parameters, and models is

$$P(\boldsymbol{x}, h, \boldsymbol{\sigma}, M) = L(\boldsymbol{x}|h, \boldsymbol{\sigma}, M)\pi(h, \boldsymbol{\sigma}|M)\Pi(M). \tag{1}$$

According this hierarchical model, firstly, M is sampled from $\Pi(M)$; then, the model parameters h and $\boldsymbol{\sigma}$ are sampled from $\pi(h, \boldsymbol{\sigma}|M)$; eventually, the data $\boldsymbol{x}$ are sampled from $L(\boldsymbol{x}|h, \boldsymbol{\sigma}, M)$.

Through conditioning and marginalization, $P(\boldsymbol{x}, h, \boldsymbol{\sigma}, M)$ can be used to obtain the post-data distributions of interest. By conditioning (1) on $\boldsymbol{x}$ and M, one gets the post-data probability distribution of the parameters given the model and data,

$$P(h, \boldsymbol{\sigma}|\boldsymbol{x}, M) = \frac{L(\boldsymbol{x}|h, \boldsymbol{\sigma}, M)\pi(h, \boldsymbol{\sigma}|M)}{Z(\boldsymbol{x}|M)}, \tag{2}$$

where the normalizing factor

$$Z(\boldsymbol{x}|M) = \int_{-\infty}^{+\infty} \mathrm{d}h \int_{\Sigma} L(\boldsymbol{x}|h, \boldsymbol{\sigma}, M)\pi(h, \boldsymbol{\sigma}|M)\, \mathrm{d}\boldsymbol{\sigma} \tag{3}$$