**THE EUROPEAN
PHYSICAL JOURNAL B**

Bayesian inference of the mean power of several Gaussian data

Giovanni Mana[1,2,a] and Carlo Palmisano[3,b]

[1] INRIM-Istituto Nazionale di Ricerca Metrologica, Strada delle cacce 91, 10135 Turin, Italy
[2] Dipartimento di Fisica, UNITO-Università di Torino, via Pietro Giuria 1, 10125 Turin, Italy
[3] DMA-Diagnostic Monitoring Applications, Turin, Italy

Abstract. The uniform prior probability density for the means of normal data leads to inconsistent Bayesian inference of their mean power and jeopardizes the possibility of selecting among different models that explain the data. We reinvestigated the problem avoiding delivering unrecognised information and looking at it in a novel way. Namely, to consider a finite power, we used a normal prior minimally diverging from the uniform one, hyperparameterised by the mean and variance, and left the data to choose the most supported parameters. We also obtained an extended James–Stein estimator averaging the hyperparameters and avoiding empirical Bayes techniques.

1 Introduction

The Bayes theorem encodes the measurement uncertainty in the probability assignments to the possible values of a measurand. It also applies to encoding the lack of certainty of the data model. If the data explanations are questionable and different models compete, their posterior probabilities provide the framework to assess the model uncertainties. The model likeliness is proportional to the data evidence (or marginal likelihood). Since the evidence must remain unchanged after changes in the model parameters, the prior distributions of different parameterisations must be proper and comply with the change-of-variable rule.

When no objective prior information is available, subjective inferences provide the required compliance with the transformation of the prior distributions under parameter changes. Alternatives, discussed by [16] and known as Jeffreys' priors, are distributions obtained as a model functional encoding the symmetries of the measurand link with the data. Others, developed by [5] and [3], maximise the mutual information between the data and measurand, which measures how much we learn about the measurand from the data.

If improper, these distributions can lead to inconsistencies. For example, the Jeffreys' uniform prior over the reals for the mean of Gaussian data leads

to an inconsistent inference of their mean power, see [4,5,12,13,28].

We consider the simultaneous inferences of the individual means of multivariate Gaussian data, individual means' squares, and average means' squares. Supposing the data are the measured values of a discrete-time signal, we will refer to the lasts as powers. The uniform distribution for the means' prior delivers information about their squares. To avoid introducing this information into the problem and taking a finite power into account, we investigate a novel solution which extends the approaches of [1,8,9,24,25].

Our investigations were prompted by related problems in the estimate of the power of signals and the modulus of complex quantities from measurement results affected by additive uncorrelated Gaussian errors, which were lengthy discussed in [2,6,14,15,19,32].

The manuscript is organised as follows. Section 2 states the problem, identifies its origin, and outlines a solution. Next, we overview hierarchical modelling and averaging. In Sect. 4, apply them to derive consistent inferences of the data power. Numerical examples are given in Sect. 5.

All the integrations were carried out in terms of standard mathematical functions with the aid of Mathematica [30]. The relevant notebook is given as supplementary material. To read and interact with it, download the Wolfram Player free of charge [31].

[a] e-mail: g.mana@inrim.it (corresponding author)
[b] e-mail: cpalmisano@dmatorino.it

2 Stein paradox

2.1 Problem statement

In the simplest form of the Stein paradox, $x_i \sim N(x_i|\mu_i, \sigma = 1)$ are observations of independent normal variables with unknown means μ_i and known variance σ^2, for $i = 1, 2, ..., m$. With a somewhat inconsistent use of notation, we will use the same symbol to indicate both the random variables and the labels of their possible values. To keep the algebra simple, without loss of generality, we use units where $\sigma = 1$.

For example, x_i are samples of a $\mu(t)$ signal affected by additive white Gaussian noise, where the uncertainty for each datum is known, as in [6] and [2]. Suppose that, in addition to the measurands $\mu_i = \mu(t_i)$, we are also interested in the signal power, that is to $\theta^2 = |\boldsymbol{\mu}|^2/m$, where $\boldsymbol{\mu} = \{\mu_1, \mu_2, ... \mu_m\}$ [32].

The Jeffreys prior of every instantaneous signal μ_i is the uniform distribution over the reals $U_\infty(\mu_i) \propto$ const., resulting in independent normal posteriors, $\mu_i \sim N(\mu_i|x_i, \sigma = 1)$. By changing variable in the posterior distributions, μ_i^2 are independent non-central χ_1^2 variables having one degree of freedom, mean $1 + x_i^2$, and variance $2(1 + 2x_i^2)$. Similarly, $m\theta^2$ is a non-central χ_m^2 variable having m degrees of freedom. Therefore, it follows that the posterior mean and variance of θ^2 are

$$\mathbb{E}(\theta^2|\mathbf{x}, \mathscr{M}_\infty) = 1 + \overline{x^2}, \tag{1a}$$

where $\mathbf{x} = \{x_1, x_2, ... x_m\}$, $\overline{x^2} = |\mathbf{x}|^2/m$, and $\mathscr{M}_\infty$ is the data model assuming the uniform prior, and

$$\mathrm{Var}(\theta^2|\mathbf{x}, \mathscr{M}_\infty) = \frac{2}{m}\left(1 + 2\overline{x^2}\right). \tag{1b}$$

From a frequentist viewpoint, since $m\overline{x^2}$ is a noncentral χ_m^2 variable having m degrees of freedom, $\overline{x^2}$ is a biased estimator of θ^2. In fact,

$$\mathbb{E}(\overline{x^2}|\boldsymbol{\mu}) = 1 + \theta^2 \tag{2a}$$

and

$$\mathrm{Var}(\overline{x^2}|\boldsymbol{\mu}) = \frac{2}{m}\left(1 + 2\theta^2\right). \tag{2b}$$

As m tends to infinity, provided that, θ^2 and, consequently, $\overline{x^2}$ converge, a bad situation occurs: (1a) and (1b) jointly predict that θ^2 is certainly $\overline{x^2} + 1$, but, at the same time, (2a) and (2b) jointly predict that θ^2 is certainly $\overline{x^2} - 1$.

2.2 Paradox explanation

The $U_\infty(\mu_i) \propto$ const. usage can be justified by showing that the $\mu_i \sim N(\mu_i|x_i, \sigma = 1)$ posterior is a suitable limit obtained from proper priors having increasingly large variance [3]. The difference between the Bayesian

and frequentist certainties is due to the information encoded in $U_\infty(\mu_i)$, which makes the data explanations underlying the Bayesian and frequentist analyses different.

As most of its mass is at infinite, $U_\infty(\mu_i)$ encodes that $|\mu_i|$ is greater than any positive number. This information is irrelevant to μ_i; but, it is not for μ_i^2. It is worth noting that, if really $\mu_i \sim U_\infty(\mu_i)$, then $\overline{x^2}$ and θ^2 must diverge and the inconsistency disappears. Therefore, the problem originates in the conflict between the information encapsulated in $U_\infty(\mu_i)$ and the data, see also [11]. Also, if we are not happy with the difference between (1a) and (2a), it means that we believe that θ^2 and, consequently, μ_i^2 are finite.

2.3 Proposed solution

To remove the inconsistency between (1a) and (2a), we must take the assumption that $|\mu_i|$ is bounded into account and use a proper prior. Therefore, extending [1], we set $\mu_i|a_i, b_i \sim N(\mu_i|a_i, b_i)$, where the mean and standard deviation a_i and b_i are hyper-parameters and the $b_i \to \infty$ limit is the uniform distribution. The normal distribution has been chosen because it has the minimum relative entropy concerning the uniform one under a fixed variance [26,27].

The simplest way to encode $|\mu_i| < \infty$ is setting $a_i = a$ and $b_i = b$ for all the samples, which condition is sufficient, though not necessary, for the convergence. If the samples are indistinguishable, i.e., the data labelling is unknown, assigning the same mean and variance is reasonable.

Hence, we assume the prior

$$\boldsymbol{\mu}|a, b \sim \pi(\boldsymbol{\mu}|a, b) = \prod_{i=1}^{m} N(\mu_i|a, b). \tag{3}$$

According to the Bayesian viewpoint, firstly, the $\boldsymbol{\mu}$ measurands are sampled from the $\pi(\boldsymbol{\mu}|a, b)$ prior, then the $\mathbf{x}$ data are sampled from the $N(\mathbf{x}|\boldsymbol{\mu}, \sigma = 1)$ distribution. Therefore, different priors identify different models, and we can let the data select the most likely.

Since the prior (3) encodes the belief that all the measurands (e.g., the instantaneous signals μ_i) have the same mean, it will originate a shrinkage of the means' inferences on the sample mean. Our prior choice is not new. However, previous investigations, for instance, [4,12,24], set $\mu_i|b \sim N(\mu_i|0, b)$, which encodes the strongest belief that all the measurands are expected to have a zero mean.

3 Outline of the Bayesian model selection

To determine the most likely prior, we let the data choose. Let $N(\mathbf{x}|\boldsymbol{\mu}, \sigma = 1)$ be the multivariate distribution of the $\mathbf{x}$ data. Hence, the hierarchical models

competing to explain the data are

$$\mathcal{M}_{ab} = \{N(\mathbf{x}|\boldsymbol{\mu}, \sigma = 1) : \boldsymbol{\mu} \sim \pi(\boldsymbol{\mu}|a,b)\}, \qquad (4)$$

where $\pi(\boldsymbol{\mu}|a,b)$ is the prior distribution (3) and a and b index the models. The posterior distribution of the measurands is

$$\Pi(\boldsymbol{\mu}|\mathbf{x},a,b) = \frac{N(\mathbf{x}|\boldsymbol{\mu},\sigma=1)\pi(\boldsymbol{\mu}|a,b)}{Z(\mathbf{x}|a,b)}. \qquad (5)$$

The marginal likelihood or evidence,

$$Z(\mathbf{x}|a,b) = \int_{\mathbb{R}^m} N(\mathbf{x}|\boldsymbol{\mu},\sigma=1)\pi(\boldsymbol{\mu}|a,b)\,\mathrm{d}\boldsymbol{\mu}, \qquad (6)$$

is the sampling distribution $\mathbf{x}$ given the model indexed by a and b, but no matter what the values of $\boldsymbol{\mu}$ – or of others model parameterisations – might be.

The marginal likelihood can be used to compare the competing models by their probability $Q(a,b|\mathbf{x})$ as provided by the data,

$$Q(a,b|\mathbf{x}) = \frac{Z(\mathbf{x}|a,b)\varpi(a,b)}{\displaystyle\int_0^\infty \int_{-\infty}^{+\infty} Z(\mathbf{x}|a',b')\varpi(a',b')\,\mathrm{d}a'\,\mathrm{d}b'}, \qquad (7)$$

where $\varpi(a,b)$ is the prior probability of a,b and the integration is carried out on its support.

The model-averaged posterior of $\boldsymbol{\mu}$ is

$$P(\boldsymbol{\mu}|\mathbf{x}) = \int_0^\infty \int_{-\infty}^{+\infty} \Pi(\boldsymbol{\mu}|\mathbf{x},a,b)Q(a,b|\mathbf{x})\,\mathrm{d}a\,\mathrm{d}b, \quad (8)$$

which can also be obtained by marginalising (5) for the hyper-parameters. The model uncertainty can be embedded in the expected value of the data mean by

$$\begin{aligned}
\mathbb{E}(\boldsymbol{\mu}|\mathbf{x}) &= \int_{\mathbb{R}^m} \boldsymbol{\mu} P(\boldsymbol{\mu}|\mathbf{x})\,\mathrm{d}\boldsymbol{\mu} \\
&= \int_0^\infty \int_{-\infty}^{+\infty} \mathbb{E}(\boldsymbol{\mu}|\mathbf{x},a,b)Q(a,b|\mathbf{x})\,\mathrm{d}a\,\mathrm{d}b, \quad (9)
\end{aligned}$$

where

$$\mathbb{E}(\boldsymbol{\mu}|\mathbf{x},a,b) = \int_{\mathbb{R}^m} \boldsymbol{\mu}\,\Pi(\boldsymbol{\mu}|\mathbf{x},a,b)\,\mathrm{d}\boldsymbol{\mu}. \qquad (10)$$

4 Application to the Stein paradox

4.1 Posteriors of the instantaneous signals

By application of (5) and (6), the prior (3) results in independent and identically distributed μ_i having marginal likelihood

$$Z(x_i|a,b) = \frac{1}{\sqrt{2\pi(1+b^2)}} \exp\left[-\frac{(x_i-a)^2}{2(1+b^2)}\right] \quad (11)$$

and normal posterior

$$\Pi(\mu_i|x_i,a,b) = N(\mu_i|\overline{\mu_i},\sigma_\mu), \qquad (12)$$

where

$$\overline{\mu_i} = \mathbb{E}(\mu_i|x_i,a,b) = \frac{a+b^2 x_i}{1+b^2} \qquad (13a)$$

and

$$\sigma_\mu^2 = \mathrm{Var}(\mu_i|a,b) = \frac{b^2}{1+b^2} \qquad (13b)$$

are the posterior mean and variance of μ_i. It is worth noting that if $b \to \infty$ then $\overline{\mu_i} = x_i$ and $\sigma_\mu^2 = 1$. The relevant integrations are given in the supplementary material.

4.2 Hyper-prior

We assign prior probabilities to a and b to continue the analysis. The sampling distribution of $\mathbf{x}$ given a and b is (see the supplementary material)

$$Z(\mathbf{x}|a,b) = \prod_{i=1}^m Z(x_i|a,b) = \frac{\exp\left\{-\dfrac{m\left[s_x^2 + (\overline{x}-a)^2\right]}{2(1+b^2)}\right\}}{\sqrt{(2\pi)^m(1+b^2)^m}}, \qquad (14)$$

where $\overline{x} = \sum_{i=1}^m x_i/m$ and $s_x^2 = \sum_{i=1}^m (x_i-\overline{x})^2/m$ are the sample mean and (biased) sample variance, respectively.

Since any of the models (4) is uncertain, to offer evidence or disprove that it explains the data, they must allow for comparisons. This requires proper prior distributions of different parameterisations and compliance with the change-of-variable rule. In the absence of measurable information, the normalised Jeffreys' hyperprior (see the supplementary material),

$$\varpi(a,b) = \frac{b}{V_a\sqrt{(1+b^2)^3}}, \qquad (15)$$

where $b > 0$ and V_a is the length of the a's domain, does the work. It is worth noting that (15) preserves the convergence of θ^2 when $m \to \infty$.

4.3 Model probabilities

By application of (7) with (14) and (15), the probability of a,b explaining the data is (see the supplementary material)

$$Q(a,b|\overline{x},s_x^2) = \frac{\sqrt{2m}\,u_x^m b}{\sqrt{\pi(1+b^2)^{m+3}}\,\Gamma(m/2,0,u_x^2)}$$
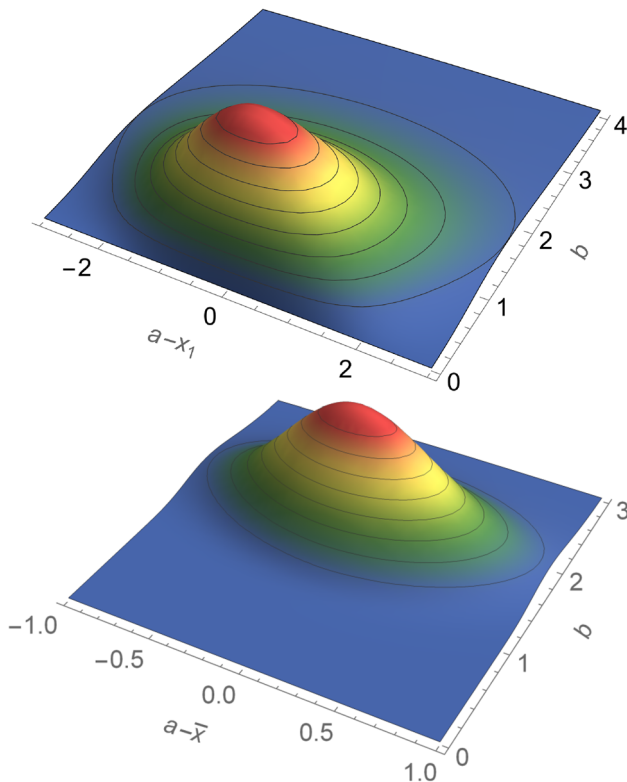
Fig. 1 Posterior probability densities $Q(a, b|\overline{x}, s_x^2)$ of the model $\mathcal{M}_{ab}$ given the $\{x_1, x_2, \ldots x_m\}$ normal data. Top: $m = 1$. Bottom: $m = 20$ and $s_x = 2$

$$\times \exp\left\{-\frac{m\left[s_x^2 + (\overline{x} - a)^2\right]}{2(1 + b^2)}\right\}, \quad (16)$$

where $u_x^2 = ms_x^2/2$ and $\Gamma(a, z_1, z_2)$ is the generalized incomplete gamma function [20]. Since V_a simplifies in the $Q(a, b|\overline{x}, s_x^2)$ calculation and provided it is large enough to approximate (7) by extending the integration over a to the reals, there is no need to introduce additional undefined parameters.

Figure 1 shows (16) when $m = 1$ (top) and $m = 20$ and $s_x = 2$ (bottom). It is worth noting that $Q(a, b|\overline{x}, s_x^2)$ depends only on the sample mean and biased variance; its mode is $a = \overline{x}$ and, as $m \to \infty$, $b^2 = s_x^2 - 1$ (see the supplementary material).

4.4 Expectations of the instantaneous signals

4.4.1 $m = 1$ case

By application of (7) with (11) and (15), the probability density reduces to

$$Q(a, b|x_1) = \frac{b}{\sqrt{2\pi}(1 + b^2)^2} \exp\left[-\frac{(x_1 - a)^2}{2(1 + b^2)}\right], \quad (17)$$

which can also be obtained as the $s_x^2 \to 0$ limit of (16) evaluated for $m = 1$ (see the supplementary material).

The most supported model is indexed by $a = x_1$ and $b = 1/\sqrt{3}$, whereas the $b \to \infty$ model – corresponding to the uniform prior – is excluded.

According to (8), after averaging the μ_1 posterior (12) over the model probability (17),

$$\mu_1|x_1 \sim P(\mu_1|x_1) = \int_0^\infty \int_{-\infty}^{+\infty} N(\mu_1|\overline{\mu_1}, \sigma_\mu)Q(a, b|x_1)\, \mathrm{d}a\, \mathrm{d}b$$
$$= N(\mu_1|x_1, \sigma = 1), \quad (18)$$

i.e. the distribution of μ_1 is a normal distribution having mean x_1 and unit variance, and the distribution of μ_1^2 is a non-central χ_1^2 distribution, having one degree of freedom and non-centrality parameter x_1^2 (see the supplementary material). These are important and non-trivial results. They demonstrate that the hierarchical models (4) are consistent with the uniform prior (see Sect. 2.1) and that there is no hyper-prior effect on the posterior distributions of μ_1 and μ_1^2, as it occurs in [4].

4.4.2 $m \geq 2$ case

By substitution of (13a) and (16) into (9), the model-averaged value of $\mathbb{E}(\mu_i|x_i, a, b)$ is (see the supplementary material)

$$\mathbb{E}(\mu_i|x_i, \overline{x}, s_x^2) = \overline{x} + (1 - R)(x_i - \overline{x}), \quad (19a)$$

where

$$R = \frac{\Gamma(m/2 + 1, 0, u_x^2)}{\Gamma(m/2, 0, u_x^2)u_x^2}, \quad (19b)$$

$u_x^2 = ms_x^2/2$ and $\Gamma(a, z_1, z_2)$ is the generalised incomplete gamma function [20]. Given the belief encoded in the priors, this inference minimises the (Bayesian) quadratic risk. It belongs to the estimator class considered in [21], but it does not comply with the condition required to dominate (from a frequentist perspective) the James–Stein estimator. In this regard, we note that this paper is not about the dominance over the James–Stein estimators, but about highlighting and encoding the belief $\theta^2 < \infty$.

If $m = 1$ then $s_x = 0$ and $\overline{x} = x_i$. In this case, the last term of (19a) vanishes and the mean is the observed value (see the $m = 1$ case). Also, since its value is irrelevant, we set $R(m = 1)$ conventionally to zero (incidentally, the $s_x \to 0$ limit is $1/3$). As shown in Fig. 2, the mean (19a) is between x_i and the sample mean $\overline{x}$. This behaviour follows from the mild assumption of a constant μ_i (see Sect. 2.3) encapsulated in the most supported data model.

As shown in the supplementary material, when $s_x^2 \ll 1$, (19a) is approximated by

$$\mathbb{E}(\mu_i|x_i, \overline{x}, s_x^2 \ll 1) \approx \overline{x}$$
$$+ \left(1 - \frac{m}{m + 2 - m^2 s_x^2/2}\right)(x_i - \overline{x}), \quad (20a)$$
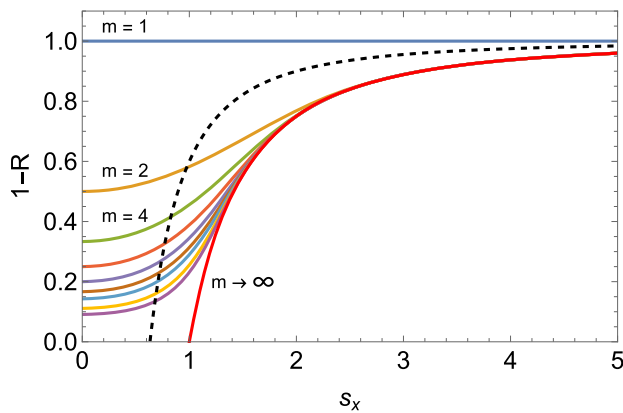
Fig. 2 Shrinking factor of the model-averaged posterior mean of μ_i *vs* the sample standard-deviation. Nine cases are considered, $m = 1$ (top line), 2, 4, 6, 8, 10, 12, 16, 20 (bottom line). The dashed line is the James–Stein estimate (21), when $m = 5$. The red line is the $m \to \infty$ limit of both the model-averaged mean and James–Stein estimate; $s_x < 1$ is meaningless in this case, see the supplementary material

which, as the sample size increases and $s_x \to 0$, tends to $\overline{x}$. When $s_x^2 \gg 1$, (19a) is approximated by

$$\mathbb{E}(\mu_i|x_i, \overline{x}, s_x^2 \gg 1) \approx \overline{x} + \left(1 - \frac{1}{s_x^2}\right)(x_i - \overline{x}) \quad (20b)$$

and supports the x_i datum. For many observations, we obtain

$$\lim_{m \to \infty} \mathbb{E}(\mu_i|x_i, \overline{x}, s_x^2) = \overline{x} + \left(1 - \frac{1}{s_x^2}\right)(x_i - \overline{x}). \quad (20c)$$

As shown in the supplementary material, when $m \to \infty$, it is certain that $s_x^2 \geq 1$. These asymptotic expressions are consistent with the expectation that a sample variance larger than the data variance (which was set to one) supports a varying signal and a smaller one the opposite.

4.5 James–Stein estimate

Empirical Bayes methods set the a and b hyperparameters in (13a) to specific values, see [1], instead of integrating them out. For instance, if in (13a) and following [8] we set a to its posterior mode $\overline{x}$ and $1 + b^2$ to $ms_x^2/(m-3)$, then $\overline{\mu_i}$ reduces to the (positive) James–Stein estimate given in [9,10],

$$\mu_i^{\text{JS}} = \begin{cases} \overline{x} + \left(1 - \frac{m-3}{ms_x^2}\right)(x_i - \overline{x}) & \text{if } s_x^2 \geq \frac{m-3}{m} \\ \overline{x} & \text{if } s_x^2 \leq \frac{m-3}{m} \end{cases}, \quad (21)$$

the first line of which is derived in the supplementary material and shown in Fig. 2 for $m = 5$. The replacement in the second line avoids pulling the estimate away from the $[x_i, \overline{x}]$ interval, see [10]. The reason for the $1 + b^2 = ms_x^2/(m-3)$ choice resides in the fact that

$(m-3)/(ms_x^2)$ is an unbiased estimator of $1/(1 + b^2)$, see [8].

In Sect. 4.3 we have shown that, as $m \to \infty$, the b^2 mode is $s_x^2 - 1$, see (16). By using this value in (13a), we obtain (see the supplementary material) the same asymptotic limit of (21).

4.6 Expectations of the instantaneous powers

From (12), it follows that the normalized powers $(\mu_i/\sigma_\mu)^2$ are independent non-central χ_1^2 variables having one degree of freedom and non-centrality parameter $\lambda_i = (\overline{\mu_i}/\sigma_\mu)^2$, where $\overline{\mu_i}$ and σ_μ^2 are given by (13a) and (13b), respectively. Hence,

$$\frac{\mu_i^2}{\sigma_\mu^2}\bigg|x_i, a, b \approx \chi_1^2(\mu_i^2/\sigma_\mu^2|\lambda_i). \quad (22)$$

By taking the mean of the non-central χ_1^2 distribution and the μ_i/σ_μ normalisation into account, the posterior means of the μ_i^2 powers is

$$\mathbb{E}(\mu_i^2|x_i, a, b) = \sigma_\mu^2 + \overline{\mu_i}^2$$
$$= \frac{a^2 + b^2(1 + 2ax_i) + b^4(1 + x_i^2)}{(1 + b^2)^2} \quad (23)$$

and, by application of (9) to (23), its model-averaged value is

$$\mathbb{E}(\mu_i^2|x_i, \overline{x}, s_x^2) = x_i^2 + 1 + S, \quad (24a)$$

where (see the supplementary material)

$$S = \frac{1}{\Gamma(m/2, 0, u_x^2)u_x^4}\bigg\{\Gamma(2 + m/2, 0, u_x^2)(\overline{x} - x_i)^2$$
$$+ \Gamma(1 + m/2, 0, u_x^2)[1/m - 1 + 2x_i(\overline{x} - x_i)]u_x^2\bigg\}, \quad (24b)$$

$u_x^2 = ms_x^2/2$, and $\Gamma(a, z_1, z_2)$ is the generalised incomplete gamma function [20].

The asymptotic means of the μ_i^2 powers are derived in the supplementary material. When $s_x^2 \ll 1$ the data support equal μ_i,

$$\mathbb{E}(\mu_i^2|x_i, \overline{x}, s_x^2 \ll 1) \approx \overline{x}^2 + \frac{6 - m^2 s_x^2}{2(m+2) - m^2 s_x^2}, \quad (25a)$$

where we used $x_i \to \overline{x}$, and the mean shrinks to $\overline{x}^2$. If $m = 1$, then $s_x = 0$ and the mean is $x_1^2 + 1$. A large sample variance supports different μ_i and

$$\mathbb{E}(\mu_i^2|x_i, \overline{x}, s_x^2 \gg 1) \approx x_i^2 + 1 - \frac{m - 2m(\overline{x} - x_i)x_i - 1}{ms_x^2}. \quad (25b)$$
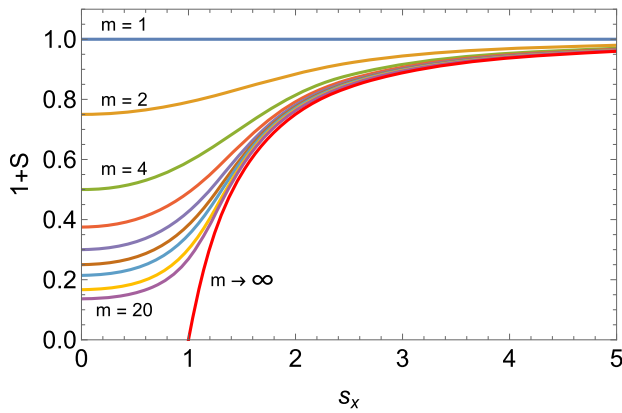
Fig. 3 Offset of the model-averaged posterior mean of μ_i^2 vs the sample standard-deviation, when $x_i = \overline{x}$. Nine cases are considered, $m = 1$ (top line), 2, 4, 6, 8, 10, 12, 16, 20 (bottom line). The red line is the $m \to \infty$ limit; $s_x < 1$ is meaningless in this case, see the supplementary material
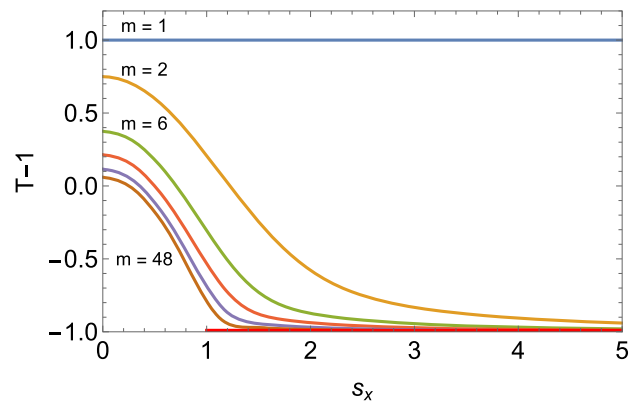

Fig. 4 Offset of the model-averaged expectation of θ^2 vs the sample standard-deviation. Six cases were considered, $m = 1$ (top line), 2, 6, 12, 24, 48 (bottom line). The red line is the $m \to \infty$ limit; $s_x < 1$ is meaningless in this case, see the supplementary material

Eventually, when m tends to the infinity,

$$\lim_{m \to \infty} \mathbb{E}(\mu_i^2 | x_i, \overline{x}, s_x^2) = x_i^2$$
$$+ 1 + \frac{(\overline{x} - x_i)^2}{s_x^4} - \frac{1 - 2(\overline{x} - x_i)x_i}{s_x^2}, \quad (25c)$$

where $s_x^2 \geq 1$, see the supplementary material. Since x_i, $\overline{x}$, and s_x^2 are not independent, a general graphical display of (24a) is impossible. To make it feasible, in Fig. 3, we considered the $x_i = \overline{x}$ case.

4.7 Expectation of the mean power

Let us turn the attention to the mean power $\theta^2 = |\boldsymbol{\mu}|^2/m$. From (12), $m\theta^2/\sigma_\mu^2$ is a non-central χ_m^2 variable having m degrees of freedom and non-centrality parameter $\lambda = \sum(\overline{\mu_i}^2/\sigma_\mu^2)$, where $\overline{\mu_i}^2$ and σ_μ^2 are given by (13a) and (13b), respectively. Hence, from (23), the expectation and variance of the mean power are (see the supplementary material)

$$\mathbb{E}(\theta^2 | \overline{x}, s_x^2, a, b) = \frac{(m + \lambda)\sigma_\mu^2}{m}$$
$$= \frac{a^2 + b^2(1 + 2a\overline{x}) + b^4(1 + s_x^2 + \overline{x}^2)}{(1 + b^2)^2}$$
$$(26a)$$

and

$$\mathrm{Var}(\theta^2 | \overline{x}, s_x^2, a, b) = \frac{2(m + 2\lambda)\sigma_\mu^4}{m^2}$$
$$= \frac{2b^2 \left[2a^2 + b^2(1 + 4a\overline{x}) + b^4(1 + 2s_x^2 + 2\overline{x}^2) \right]}{m(1 + b^2)^3},$$
$$(26b)$$

Averaging $\mathbb{E}(\theta^2 | \overline{x}, s_x^2, a, b)$ over the models via (9), we obtain (see the supplementary material)

$$\mathbb{E}(\theta^2 | \overline{x^2}, s_x^2) = \overline{x^2} - 1 + T, \quad (27a)$$

where $\overline{x^2} = |\mathbf{x}|^2/m$ is the sample mean power,

$$T = \frac{3m\Gamma(m/2, 0, u_x^2) + 2(2u_x^2 - 3)u_x^m e^{-u_x^2}}{2mu_x^2\Gamma(m/2, 0, u_x^2)}, \quad (27b)$$

$u_x^2 = ms_x^2/2$, and $\Gamma(a, z_1, z_2)$ is the generalised incomplete gamma function [20]. This inference, which minimises the (Bayesian) quadratic risk, belongs to estimator classes previously considered by [7,17,18,23–25].

As shown in Fig. 4, when $s_x^2 \ll 1$ the data support $\mu_i = \mathrm{const.}$ and

$$\mathbb{E}(\theta^2 | \overline{x^2}, s_x^2 \ll 1) \approx \overline{x^2} - 1$$
$$+ \frac{m + 5}{m + 2} - \frac{m(14 + 6m + m^2)s_x^2}{(m + 2)^2(m + 4)}. \quad (28a)$$

If $m = 1$, then $s_x = 0$ and the power of this datum is again $x_1^2 + 1$. When $s_x^2 \gg 1$, the data support a varying signal and

$$\mathbb{E}(\theta^2 | \overline{x^2}, s_x^2 \gg 1) \approx \overline{x^2} - 1 + \frac{3}{ms_x^2}. \quad (28b)$$

Eventually, it is non-obvious and remarkable that, as $m \to \infty$, the expectation of the mean power converges to the frequentist estimate $\overline{x^2} - 1$, see (2a). In fact,

$$\mathbb{E}(\theta^2 | \overline{x^2}, s_x^2, m \gg 1) \approx \overline{x^2} - 1 + \frac{3}{ms_x^2}, \quad (28c)$$

where $s_x^2 \geq 1$, see the supplementary material. These asymptotic expressions are derived in the supplementary material.
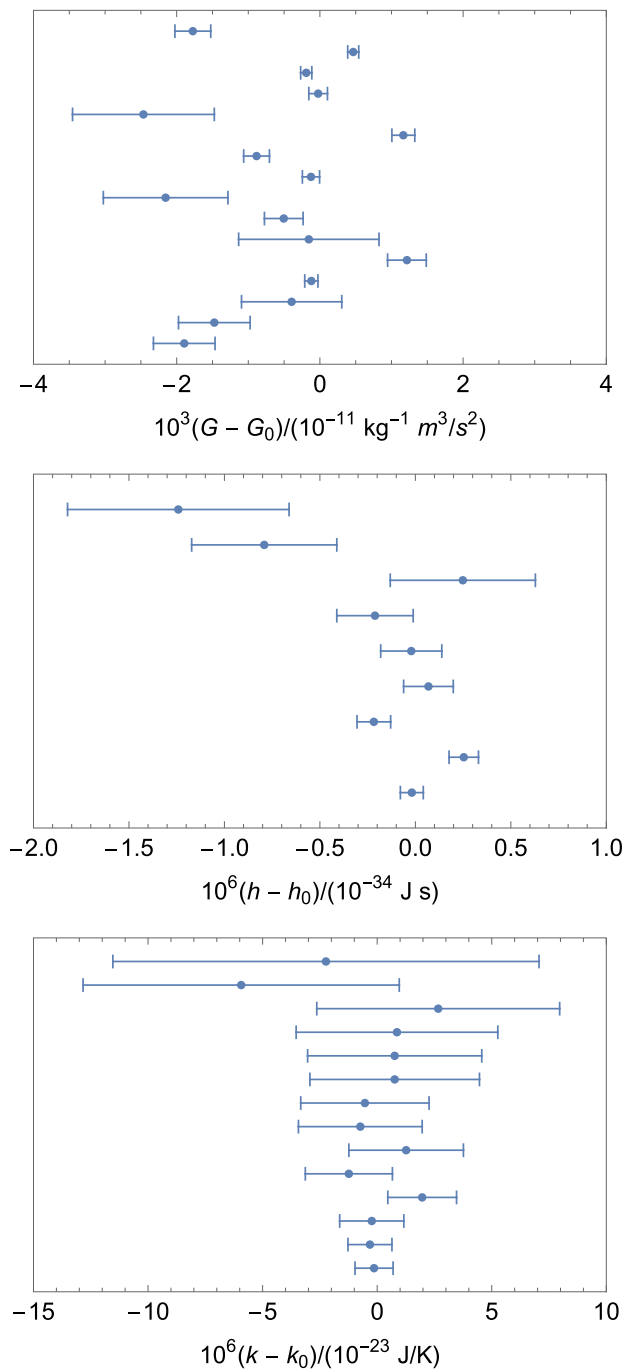
Fig. 5 Differences of the measured values of the Newtonian constant of gravitation G, the Planck constant h, and the Boltzmann constant k given in [22] and [29] from their weighted mean, G_0, h_0, and k_0, respectively

5 Application examples

According to the Bayes theorem, the posterior probability of a model is proportional to the marginal likelihood of its parameters based on the data. However, if the parameter prior-density is improper, the marginal likelihood can not be determined. In fact, when a probability density is non-integrable, it is given only up to an arbitrary scale factor, which means that the marginal likelihood depends on the chosen value of this factor. This is the case of the Jeffreys' uniform prior over the reals for the mean of Gaussian data.

The problem is evaded by the prior (3), which has been proved to produce sound posteriors for the data mean and power, while avoiding inconsistencies, and, contrary to the uniform one, is proper and encodes a finite measurand value.

To give examples, we considered the measured values of the Newtonian constant of gravitation G, the Planck constant h, and the Boltzmann constant k given in [22] and [29]. These measured values have been used by the CODATA Task Group on Fundamental Physical Constants to determine mutually consistent values for use in science and technology [29]. Their differences from the weighted mean are shown in Fig. 5.

These examples have been selected to represent the cases where a visual inspection of the data suggests disagreement (G values), agreement (k values), or uncertain judgment (h values). Where the data are mutually inconsistent, most probably they reflect systematic errors. Still, it is possible that they are pointing to unknown subtleties, – perhaps the constant value depend on how it is measured.

The objective of a Bayesian equal-mean test is to quantify these qualitative judgments by assigning them probabilities. Therefore, we compare the hypothesis H_0 that the measured values are sampled from Gaussian distributions having the same mean against that they are sampled from Gaussian distributions whose means might be different, H_1. Assuming the same 50% prior probability of the two data models, their posterior probabilities are

$$\text{Prob}(H_n|\mathbf{x}) = \frac{Z(\mathbf{x}|H_n)}{Z(\mathbf{x}|H_1) + Z(\mathbf{x}|H_2)}, \qquad (29)$$

where $Z(\mathbf{x}|H_n)$ is the marginal likelihood of the n-th model parameters.

Calculating the $Z(\mathbf{x}|H_1)$ marginal likelihood in the simplest way, by resting on the previous results, requires equal and unit variances of the input data. The unequal variance case makes the algebra cumbersome without adding conceptual news. To comply with the unit variances constraint, we consider the normalised differences $(x_i - x_0)/u_i$ of the measured values from their weighted mean x_0, where u_i^2 is the variance of the i-th datum. However, these scaled data only have the same mean if it equals x_0. Therefore, we must restrict H_0 to this case and, to take the x_0 variance, σ_0^2, into account, increase the data variances to $\sigma_i^2 = u_i^2 + \sigma_0^2$.

5.1 H_0 hypothesis

Let us consider the normalised differences $(x_i - x_0)/\sigma_i$ of the measured values from their weighted mean x_0, where $\sigma_i^2 = u_i^2 + \sigma_0^2$ is the sum of variances of the i-th datum and the mean, u_i^2 and σ_0^2, respectively. If each normalised difference is independently sampled from

the same Gaussian distributions having zero mean and unit variance, their joint distribution is

$$L(\mathbf{x}|H_0) = \prod_{i=1}^{m} N(x_i|\mu = 0, \sigma = 1). \quad (30)$$

Since the distribution (30) is free of parameters, the marginal likelihood coincides with it. Hence,

$$Z(\mathbf{x}|H_0) = L(\mathbf{x}|H_0) = \frac{\exp\left(-\chi^2/2\right)\exp\left[-m\overline{x}^2/2\right]}{\sqrt{(2\pi)^m}}, \quad (31)$$

where $\overline{x}$ is the arithmetic mean of the normalised data and χ^2 is the sum of the squared residuals.

5.2 H_1 hypothesis

Contrary, if the measured values are independently sampled from Gaussian distributions having (or not having) different means and standard deviations, the likelihood of the scaled data $x_i \rightarrow (x_i - x_0)/u_i$ is

$$L(\mathbf{x}|\boldsymbol{\mu}, \sigma = 1) = \prod_{i=1}^{m} \frac{\exp\left[-(x_i - \mu_i)^2/2\right]}{\sqrt{2\pi}}, \quad (32)$$

where μ_i is the scaled mean. By using the prior (3),

$$\pi(\boldsymbol{\mu}|a, b) = \prod_{i=1}^{m} \frac{\exp\left[-(\mu_i - a)^2/b^2\right]}{\sqrt{2\pi}\,b}, \quad (33)$$

the marginal likelihood $Z(\mathbf{x}|a, b)$ of the scaled data is given by (14).

To determine the most probable model in the family (4), we look for the values of the hyper-parameters a and b maximising their posterior density, $Q(a, b|\mathbf{x})$, which is given by (16). They are $a_0 = \overline{x}$ and $b_0 = \mathrm{argmax}\left[Q(a = \overline{x}, b|\mathbf{x})\right]$, which must be found numerically. Eventually,

$$Z(\mathbf{x}|H_1) = \frac{\exp\left\{-\dfrac{ms_x^2}{2(1 + b_0^2)}\right\}}{\sqrt{(2\pi)^m(1 + b_0^2)^m}}. \quad (34)$$

5.3 Results

The calculations relevant to this analysis are available in the supplementary material. The results are summarised in table 1. The posterior probabilities confirm our expectations regarding the measured G and k values and resolve the uncertainty for the h values.

The probabilities of the H_0 models are smaller than expected, which may be because we assumed not only a common mean but also that it is equal to the weighted mean of the measured values. In addition, H_1 does not exclude that the data are sampled from distributions

Table 1 Posterior probabilities of the H_0 and H_1 data model

Constant	Prob($H_0|\mathbf{x}$)	Prob($H_1|\mathbf{x}$)
G	0%	100%
h	5%	95%
k	63%	37%

with the same mean. It's worth noting that, assuming a uniform prior for the data means, the Bayesian test of equal means would have been impossible.

6 Conclusion

Given measurement results affected by additive uncorrelated Gaussian errors, we investigated the Bayesian inferences of the data means, individual means' squares, and average means' squares. The result is a new way to cope with the inconsistency originated by using a uniform prior, which inconsistency occurs because the uniform prior – contrary to what was intended and the belief that it is finite – encodes that the data power is infinite.

To minimise the difference (expressed by the Kullback–Leibler divergence) from the uniform distribution, we encoded the measurands' indistinguishability and the belief of finite measurand values in a normal prior hyper-parameterised by the mean and variance. Averaging over the unknown hyper-parameters or letting the data to chose the most supported ones removes the shortcomings of the uniform distribution.

In the case of a single datum, the inferred measurand is not biased to the smallest value, as occurs in [4], but it is the measurement result itself. With more than one datum, we derived a James–Stein estimate of every single measurand consistent with the stated belief. This result was obtained without the use of empirical methods as in [8]. We showed that, as the sample size grows, the inference of the mean power is consistent and converges to the frequentist estimate.

After proving that it produces sound posteriors for the data mean and power while avoiding inconsistencies, we applied the hyper-parameterised normal prior to determining whether the measured values of the Newtonian constant of gravitation came from populations with the same mean or not. We repeated the test using the results of the measurements of the Planck and Boltzmann constants. If we had used an improper prior for the data mean, this Bayesian test would have been impossible.

Author contributions

Giovanni Mana and Carlo Palmisano contributed equally to the manuscript.

Availability of data and material Not applicable.

Code availability A Mathematica notebook including the relevant symbolic and numerical computations is supplied as supplementary material. To read and interact with it, download the Wolfram Player free of charge from Wolfram Research.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

References

1. H. Akaike, Ignorance prior distribution of a hyperparameter and Stein's estimator. Ann. Inst. Stat. Math. **32**, 171–178 (1980)

2. F. Attivissimo, N. Giaquinto, M. Savino, A Bayesian paradox and its impact on the GUM approach to uncertainty. Measurement **45**(9), 2194–2202 (2012)

3. J.O. Berger, J.M. Bernardo, D. Sun, The formal definition of reference priors. Ann. Stat. **37**(2), 905–938 (2009)

4. J.O. Berger, J.M. Bernardo, D. Sun, Overall objective priors. Bayesian Anal. **10**(1), 189–221 (2015)

5. J.M. Bernardo, Reference analysis. In: K DD, R RC (eds) Handbook of Statistics, vol. 25 (North Holland, Amsterdam, 2005), pp. 17–99

6. C. Carobbi, Bayesian inference on a squared quantity. Measurement **48**(1), 13–20 (2014)

7. M.S. Chow, A complete class theorem for estimating a noncentrality parameter. Ann. Stat. **15**(2), 800–804 (1987)

8. B. Efron, T. Hastie, *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, 1st edn. (Cambridge University Press, Cambridge, 2016)

9. B. Efron, C. Morris, Stein's estimation rule and its competitors – an empirical Bayes approach. J. Am. Stat. Assoc. **68**(341), 117–130 (1973)

10. B. Efron, C. Morris, Data analysis using Stein's estimator and its generalizations. J. Am. Stat. Assoc. **70**(350), 311–319 (1975)

11. M. Evans, H. Moshonov, Checking for prior-data conflict. Bayesian Anal. **1**(4), 893–914 (2006). https://doi.org/10.1214/06-BA129

12. M. Evans, M. Shakhatreh, Consistency of Bayesian estimates for the sum of squared normal means with a normal prior. Sankhya A **76**(1), 25–47 (2014)

13. J.F. Ferragud, Una solucion bayesiana a la paradoja de Stein. Trabajos de Estadística e Investigación Operativa **33**, 31–46 (1982)

14. N. Giaquinto, L. Fabbiano, Examples of S1 coverage intervals with very good and very bad long-run success rate. Metrologia **53**(2), S65–S73 (2016)

15. B.D. Hall, Evaluating methods of calculating measurement uncertainty. Metrologia **45**(2), L5–L8 (2008)

16. H. Harney, *Bayesian Inference: Data Evaluation and Decisions* (Springer International Publishing, Berlin, 2018)

17. T. Kubokawa, C.P. Robert, A.K.M.E. Saleh, Estimation of noncentrality parameters. Can. J. Stat. La Revue Canadienne de Statistique **21**(1), 45–57 (1993)

18. Q. Li, J. Zhang, S. Dai, On estimating the non-centrality parameter of a chi-squared distribution. Stat. Probab. Lett. **79**(1), 98–104 (2009)

19. I. Lira, On the long-run success rate of coverage intervals. Metrologia **45**(4), L21–L23 (2008)

20. O. Marichev, M. Trott, Mathematical functions site. http://functions.wolfram.com/06.07.02.0001.01 (2019)

21. Y. Maruyama, Stein's idea and minimax admissible estimation of a multivariate normal mean. J. Multivar. Anal. **88**(2), 320–334 (2004)

22. P.J. Mohr, D.B. Newell, B.N. Taylor, E. Tiesinga, Data and analysis for the codata 2017 special fundamental constants adjustment*. Metrologia **55**(1), 125 (2018)

23. N. Neff, W.E. Strawderman, Further remarks on estimating the parameter of a noncentral chi-square distribution. Commun. Stat. Theory Methods **5**(1), 65–76 (1976). https://doi.org/10.1080/03610927608827332

24. M.D. Perlman, U.A. Rasmussen, Some remarks on estimating a noncentrality parameter. Commun. Stat. **4**(5), 455–468 (1975)

25. K.M.L. Saxena, K. Alam, Estimation of the noncentrality parameter of a chi squared distribution. Ann. Stat. **10**(3), 1012–1016 (1982). https://doi.org/10.1214/aos/1176345892

26. D. Sivia, J. Skilling, *Data Analysis: A Bayesian Tutorial* (Oxford University Press, Oxford, 2006)

27. M. Smerlak, Minimum relative entropy distributions with a large mean are gaussian. Phys. Rev. E **94**, 062107 (2016)

28. C. Stein, An example of wide discrepancy between fiducial and confidence intervals. Ann. Math. Stat. **30**(4), 877–880 (1959)

29. E. Tiesinga, P.J. Mohr, D.B. Newell, B.N. Taylor, CODATA recommended values of the fundamental physical constants: 2018*. J. Phys. Chem. Ref. Data **50**(3), 033105 (2021)

30. Wolfram Research, Inc Mathematica, Version 14.0 (Champaign, 2024a). https://www.wolfram.com/mathematica

31. Wolfram Research, Inc Wolfram Player, Version 14.0 (Champaign, 2024b). https://www.wolfram.com/player

32. G. Wuebbeler, C. Elster, On the transferability of the GUM S1 type A uncertainty. Metrologia **57**, 015005 (2020)