20th IMEKO TC4 International Symposium and
18th International Workshop on ADC Modelling and Testing
Research on Electric and Electronic Measurement for the Economic Upturn
Benevento, Italy, September 15-17, 2014

About the Frequentist and the Bayesian Approach to Uncertainty

Nicola Giaquinto[1], Laura Fabbiano[2], Amerigo Trotta[1], Gaetano Vacca[2]

[1]*Dipartimento di Ingegneria Elettrica e dell'Informazione*
[2]*Dipartimento di Meccanica, Matematica e Management*
Politecnico di Bari - Via E. Orabona 4, I-70125 Bari, Italy
[nicola.giaquinto],[laura.fabbiano], [amerigo.trotta], [gaetano.vacca] @poliba.it

Abstract – **There are two well-known and different approaches to statistical inference and hypothesis testing, i.e. the *frequentist* (or *orthodox*) and the *Bayesian* one. Consequently, there are also (if one stays in the framework of probability theory) two rival approaches to uncertainty. The present work is partly a tutorial, aimed at explaining the basic aspects of the two approaches, and their relationship with the GUM; and partly a demonstration that the implementation of the Bayesian approach in the GUM Supplement 1 is too rigid. In particular, objective Bayesianism is incompatible with the propagation of distributions prescribed in Supplement 1.**

I. INTRODUCTION

"*Basically there's only one way of doing physics but there seems to be at least two ways to do statistics, and they don't always give the same answers*." With these words, nine years ago Bradley Efron [1] has effectively summarized what is still the state of matters in statistical science. For any inference problem we have (at least) two paths, commonly known as the *frequentist*, or *orthodox approach*, and the *Bayesian approach*.

The difference between the frequentist and the Bayesian approach does not thrill the community of measurement experts. Probably, most of them have only a vague idea of the existence of the two approaches, and are inclined to think that this dispute is more philosophical than substantial. But, once again citing Efron, it can be said that "*the 250-year debate between Bayesians and frequentists is unusual among philosophical arguments in actually having important practical consequences*" [1].

This paper (which is largely based on a recent one in Italian by one of the authors [2]), first of all wants to illustrate, in the form of a small tutorial, the deep conceptual and operational differences between the two approaches. Second, it wants to show why the Bayesian approach has some advantages when making inferences, but has, on the other hand, heavy disadvantages when dealing with a core problem in uncertainty evaluation, i.e., *propagation*. The implementation of Bayesianism in the GUM and in its Supplement 1 is discussed, and shown to be too rigid. Therefore, it is suggested that in the future, the GUM:

- on one hand will include a full and correct implementation of the objective Bayesian methods;
- on the other hand, will give also procedures for evaluating the uncertainty of estimates obtained with frequentist methods, which cannot be banned or ignored.

II. A CLASSICAL EXAMPLE

We discuss a very simple example, which is substantially the same chosen by Thomas Bayes himself to explain his ideas concerning probability [3]. Let us consider an automatic welding machine, whose quality is given by its defect rate P (each welding has a probability P to be defective). In order to estimate P, $n = 100$ welds are completed, and $X = 3$ of them are defective. Question is: what is our estimate of P, and what is the uncertainty of the estimate?

A. Frequentist approach

With the frequentist (or orthodox) approach, the number of observed defective welds, X, is a random variable, while the unknown parameter P is an unknown deterministic quantity, whose probability distribution is binomial:

$$f_x(x|P) = \Pr(X = x|P) = \binom{n}{x} P^x (1-P)^{n-x} \quad (1)$$

A possible estimator of P is the maximum likelihood estimator (MLE), the value $\hat{P}$ that maximizes the probability of the observed event $X = 3$. By very simple calculations, this MLE is $\hat{P} = X/n = 0.03$, which is also an "intuitive" estimation of P. Of course, $\hat{P}$ is a function of X and is therefore itself a random variable. Its distribution, $f_{\hat{P}}(\hat{p}) = \Pr(\hat{P} = \hat{p})$, has a *known expression* (calculated by (1) by setting $\hat{p} = nx$), but *unknown values*, since they depend, like expression (1), on the unknown parameter P. The distribution of $\hat{P}$ can be represented only by assigning an hypothetical value to P: for example, Fig. 1 has been obtained by setting

ISBN-14: 978-92-990073-2-7

$P = 0.03$.

The uncertainty of the estimate $\hat{P}$ can be quantified by the standard deviation of the estimator, which is $u = \sqrt{P(1-P)/n}$. Since the parameter P is unknown, also u is uncertain and must be estimated, e.g. by replacing P with $\hat{P} = 0.03$, getting $u \cong \sqrt{\hat{P}(1-\hat{P})/n} \cong 0.0171$.

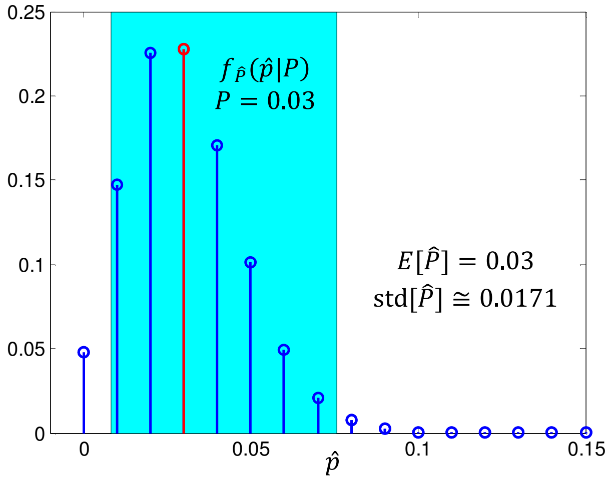

Fig. 1. Probability distribution of the frequentist estimator $\hat{P} = X/n$ in the hypothesis $P = 0.03$. The light blue area corresponds to the 90% confidence interval.

The "uncertain uncertainty" is a typical defect of the frequentist approach, but it can be circumvented by computing a suitable confidence interval which, unlike uncertainty, do not depend on unknown quantities. A 90% confidence interval for the parameter P is given by the Clopper-Pearson formula [4]

$$\mathfrak{I} = \left[B_{\frac{1-cl}{2}}(x, n-x+1);\ B_{\frac{1+cl}{2}}(x+1, n-x)\right] = [0.0082; 0.0757] \tag{2}$$

where $B_\alpha(a,b)$ is the α-th quantile of the Beta distribution with parameters a, b.

This solution to the estimation+uncertainty problem, however, is *not* acceptable according to the GUM. Neither the estimate $\hat{P}$, nor the uncertainty u, nor the confidence interval $\mathfrak{I}$ are compatible with the GUM theoretical framework.

B. Bayesian approach

Adopting the Bayesian approach, the roles of P and X are reversed. The number of defective welds, X, is a deterministic quantity, while the unknown probability P is modelled like a random variable. The problem of estimating P consists in determining its *posterior* pdf, conditioned to the observations of X, using the well-known Bayes' formula:

$$f_p(p|X) = \frac{f_X(X|p) f_p(p)}{\int f_X(X|p) f_p(p) dp} \tag{3}$$

In (3), the likelihood $f_X(X|p)$ is the probability evaluated using (1), for x equal to the observed value $X = 3$, and $f_p(p)$ is the prior distribution of P (prior, since it is assigned *before* performing the welding). If no prior information is available about P, a natural choice for $f_p(p)$ is the uniform distribution in [0;1]: the resulting posterior distribution is depicted in Fig. 2. This is a Beta distribution $B(X+1, n-X+1)$ (it can also be numerically evaluated by direct application of (3)).

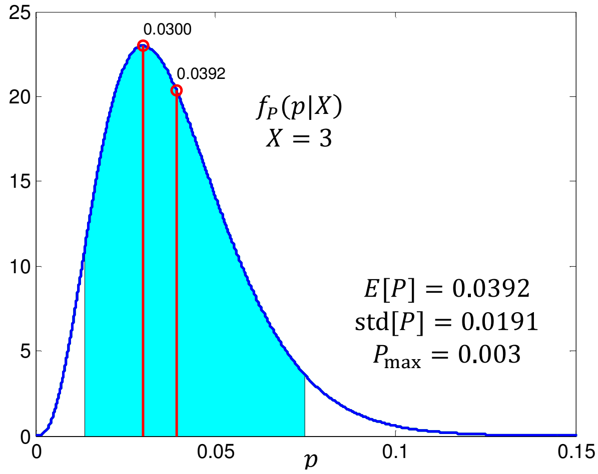

Fig. 2 – Posterior distribution of P, for a prior distribution uniform in [0;1]. The modal and mean values are highlighted, while the light blue area delimits the 90% credible interval.

Reasonable estimates of P are both the expected value $\hat{P} = E[P] = \frac{X+1}{n+2} = 0.0392$, and the modal value $\hat{P} = \text{argmax}\, f_p(p|X) = \frac{X}{n} = 0.03$. The latter choice (modal value) is tempting, because it appears logical to choose the most probable value, especially if it coincides with the intuitive frequentist estimate. However, this choice is incorrect if one wants to minimize the mean squared error (MSE) loss function, and it is also inconsistent with the uncertainty evaluation, $u = \text{std}[P] = 0.0191$, which is a measure of the dispersion of the measurand P around the expected value $E[P] = 0.0392$, not around the modal value.

If one wants to obtain the estimate $\hat{P} = 0.03$ from the Bayesian procedure, the right prior distribution is:

$$f_p(p) \propto \frac{1}{p(1-p)} \tag{4}$$

which is improper (it diverges for $p = 0$ and $p = 1$ and cannot be normalized), but yields a proper posterior distribution $B(X, n-X)$, depicted in Fig. 3.

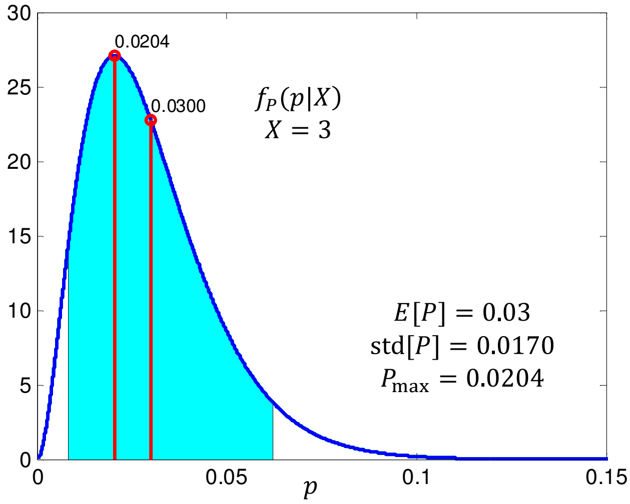

Fig. 3 – Posterior distribution of P, for a prior distribution $f_p(p) \propto \frac{1}{p(1-p)}$. The modal and mean values are highlighted, while the light blue area delimits the 90% credible interval.

With this choice of the prior, also the uncertainty of the estimate, $u = \text{std}[P] = 0.0170$, is congruent with the frequentist result.

The first advantage of the Bayesian is that, contrary to the frequentist one, the uncertainty is now *exactly known* from the available information. Besides, it is easy to evaluate an interval estimate, in the form of a *credible* or *Bayesian* interval, and to understand its meaning, also graphically. Fig. 2 and 3 show a credible interval with 90% coverage probability that, for the prior (4), and the posterior in Fig. 3, is:

$$\Im = \left[B_{\frac{1-cp}{2}}(x, n-x);\ B_{\frac{1+cp}{2}}(x, n-x)\right] = [0.0083; 0.062] \quad (5)$$

The Bayesian solution, contrary to the frequentist one, is in agreement with the GUM prescriptions. It is also very logical and, one would say, elegant. The only critical point is quite obvious: the right choice of the prior distribution. There is, however, also a second critical point, and both will be discussed in the following.

III. ASSETS AND LIABILITIES OF THE BAYESIAN APPROACH

The advantages of the Bayesian approach over the frequentist one are apparent, and the more meaningful are the following.

a) Distributions and uncertainties are exactly known without uncertainty.

b) The point estimation can be chosen according different criterions, suited to the requirements of the problem, i.e. to the definition of the loss function (in some cases the modal value might be preferable, in others the expected value, etc.).

c) It is possible to incorporate into the estimation the prior available information about the unknown quantity.

Properties (b) and (c), while expressing the power and the sophistication of the Bayesian approach, are also actually its Achilles heel. In particular, they cause problems if a *standard rule* is needed to evaluate estimate and uncertainty.

The GUM Suppement 1 prescribes that, if the only information about an unknown quantity is that it is in the interval $[a; b]$, it must be given a uniform distribution in the interval. Besides, the GUM decrees that the estimate is always the expected value (the loss function, consequently, is always the MSE). As a consequence, the GUM-compliant solution to the problem of the welding machine is $\hat{P} = \frac{X+1}{n+2} = 3.92\%$ (the expected value of the distribution in Fig. 2). But this estimate is intuitively worse than $\hat{P} = \frac{X}{n} = 3\%$ in the case of small P, and this can be easily shown mathematically.

In general, the "possibility" of incorporating into the estimation the prior available information is actually a bonding obligation: one *must* choose a prior distribution, and this choice can be very compromising. If the prior information is substantially missing, like in the example, then the prior must be chosen using some rule. This is the essence of *objective Bayesianism*, and also of the conceptual framework of the GUM and its Supplement 1. But Supplement 1 is much less sophisticated than true objective Bayesianism.

IV. THE PROBLEMS IN THE GUM IMPLEMENTATION OF BAYESIANISM

A. The prior and the posterior distribution in Bayesian theory and in the GUM

It is clear that the choice of the prior distribution in the problem of the welding machine is problematic, if one assume to know *nothing* about the value of P. In [5], five different priors are suggested, all advisable under some viewpoint. The uniform prior is in agreement with the maximum entropy principle, which is also at the basis of Supplement 1. The improper prior (4), leads to a solution consistent with the frequentist approach, and also to intuition. Another, more consistent prior is given by the Jeffreys' rule [6], which ensures the invariance of the estimate for a re-parameterization of the problem. Using the Jeffreys' rule, $f_P(p)$ must be the arcsine distribution, which gives final results a bit different from those presented above: $\hat{P} = E[P] = 0.0347$ (using the expected value), $\hat{P} = argmax\ f_P(p|X) = 0.0253$ (using the modal value). The difference is not negligible, and none of the results is close to the intuitive, and normally accepted solution, $\hat{P} = 0.03$.

Bayesian theory provides other methods to choose

$f_P(p)$, e.g. the conjugate priors, the *reference priors* by J. Bernardo, etc. [7]. A universal rule to choose $f_P(p)$ does not exist. Different methods lead to different solutions – which can be also very different, especially when few measurements are available, and the prior weighs more on the final solution.

The GUM aims exactly at such an automatic method, and this need imposes (Supplement 1, Clause 6) a quite primitive form of objective Bayesianism, in which Bayes' rule is "hidden". In Supplement 1, the unknown quantities are given directly *posterior* Bayesian distributions, on the basis of the information available on them. For example, if a quantity X is known to be in the interval $[a; b]$, it is given (as already said above) a uniform distribution in the interval; if for X is available only a measurement μ with standard uncertainty u, X is given a normal distribution $N(\mu, u^2)$; etc.

In general, giving the unknown quantity X a posterior distribution on the basis of the available information is equivalent to giving it an *improper* prior distribution $f_X(x) \propto 1$, and fixing an appropriate likelihood function. Consider the case (Supplement 1, Clause 6.4.10) of a non-negative unknown quantity X, for which the best estimate is $L = \lambda$. In this case X has (on the basis of the maximum entropy principle) the pdf

$$f_{X|L=\lambda}(x) = \frac{1}{\lambda} e^{-\lambda x} \quad (6)$$

This operation is actually equivalent to:

1) giving X the *improper* prior $f_X(x) \propto 1$;
2) describing the relationship between the measurement L and the quantity X with an exponential likelihood function $f_{L|X=x}(l) = f_{X|L=\lambda}(x)$;
3) applying the Bayes' formula (3) to obtain (6) as a posterior pdf of X, given the measurement $L = \lambda$.

It is quite clear that this simplified "Supplement 1" Bayesian approach leads always to the same distribution with the same available information: but having flexibility in choosing the distribution is a necessity in Bayesian analysis.

If one knows "nothing" about a parameter, the prior distribution must be different for different *meanings* of the parameter, even sticking to the Jeffreys' rule. If the parameter is the mean of a population, the correct prior pdf is constant, $f_X(x) \propto 1$; if it is the variance, the correct pdf is the logarithmic prior $f_X(x) \propto 1/x$; if it is the parameter λ of a Poisson distribution, the correct prior is $f_X(x) \propto \sqrt{1/\lambda}$; etc. Therefore, the first problem of the GUM method is that it is obviously too rigid in attributing a distribution on the basis of the available information on quantity.

B. The problem of the propagation of distributions

Even if the rules given in Clause 6 of Supplement 1 are not compliant with good objective Bayesianism, there is another fundamental problem:

usual propagation rules do not apply to distributions obtained with objective Bayesian methods.

To make clear this statement, we consider first an example in which the propagation of distributions does work. Consider betting on the result of the toss of a die. The first bet must be made on the value X that will result from the toss, while the second must be on the value of its square (the "output quantity" with Supplement 1 terminology): $Y = X^2$. The loss will be proportional to the square of the difference between the bet and the outcome (MSE loss function).

It is obvious that the bets that minimize the expected loss function (risk function) are $\hat{X} = E[X] = 3.5$, and $\hat{Y} = E[Y] = E[X]^2 + \text{var}[X] = 15.17$. In this case, the bet on Y (estimate of $Y = X^2$) is derived from the propagation of the distribution of X for the model $Y = f(X) = X^2$.

Now consider the same problem applied to the parameter P of the welding machine. Again, the first bet is on the value of P (the probability that one weld will be defective) and, considering the distribution of P depicted in Fig. 3, the bet is $\hat{P} = E[P] = 3\%$. The second bet is on the value of $Y = P^2$ (the probability that two consecutive welds will be both defective). By propagating the distribution, the bet on Y is

$$\hat{Y} = E[P^2] = E[P]^2 + var[P] \quad (7)$$

yielding the result $\hat{Y} = 1.53$ ‰. The distribution of Y is in Fig. 4.

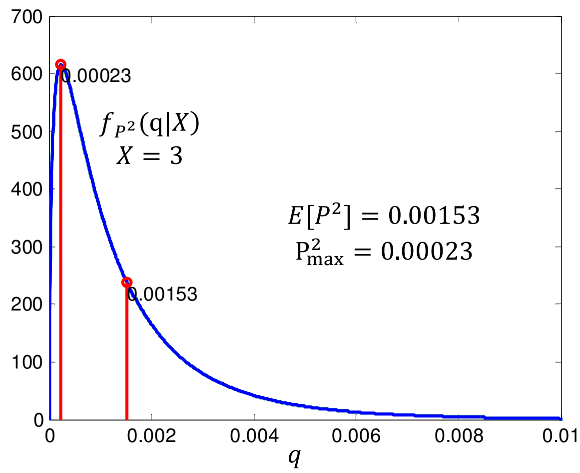

Fig. 4 – Distribution of P^2, obtained by the propagation of the pdf of P in Fig. 3. Under the MSE loss function, the estimate of P^2 is $E[P^2] = 0.00153$.

Unfortunately, even if the situation appears very similar to that of the toss of a die, there are profound differences between the distribution of X and the distribution of P, derived with an objective Bayesian method. The first distribution propagates, the second not.

The impossibility of using a simple propagation to compute the posterior pdf of P^2 is a well-known fact in the objective Bayesian methodology. For example, the posterior distribution of the mean of normal random variable cannot be propagated to estimate the square of the mean. This case is usually called "Stein's paradox" [7], [8]: observed for the first time by Stein in 1959 [9], it has been one main motivation, together with other paradoxes of objective Bayesianism [10], for the development of the theory of the "reference priors". The meaning of the paradox in the framework of the GUM has been discusses in detail in [11].

The problem of the welding machine is outside the framework of the Stein's paradox, but also in this case the propagation of the Bayesian distribution is not a good idea. We show, in particular, by using the "ugly" frequentist approach, that $\hat{Y} = E[P^2] = 1.53$ ‰ is an overestimation.

First of all, we note that the square of $\hat{P} = 0.03$ is $\hat{P}^2 = 0.9$ ‰, quite below the Bayesian estimate. Fig. 5 shows the distribution of the frequentist estimator $\hat{Y}^2 = \hat{P}^2$, computed by propagating the distribution in Fig. 1.

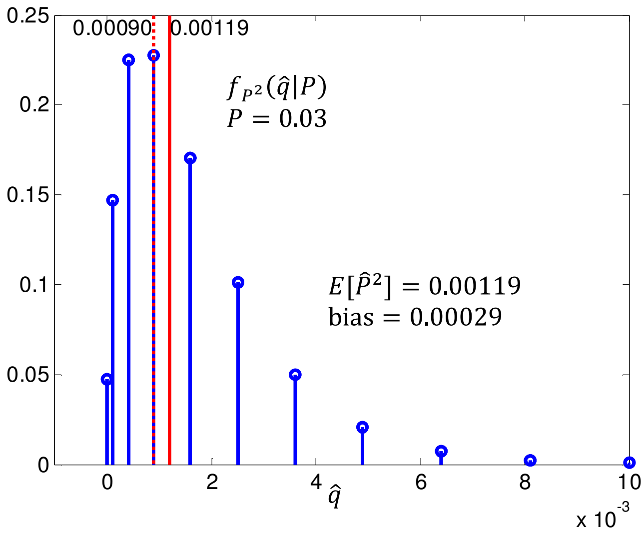

Fig. 5 – Distribution of $\hat{Y} = \hat{P}^2 = (X/n)^2$, derived by the propagation of the pdf of $\hat{P} = X/n$ in Fig. 1. The distribution shows that $\hat{P}^2$ is biased, and therefore suggests the correct estimate $\widehat{Y'} = \hat{P}^2 - bias$.

It must be remembered that distributions in Figs. 1 and 5 are computed under the hypothesis of a "true value", i.e. $P = 3\%$, $P^2 = 0.9$ ‰. The expected value of $\hat{Y}$ should be equal to P^2, but it is not, since $E[\hat{Y}] = P^2 + var[\hat{P}] > P^2$. An (approximate) unbiased estimator is given, therefore, by:

$$\widehat{Y'} = \hat{Y} - var[\hat{P}] \cong \hat{P}^2 - \hat{P}(1-\hat{P})/n \qquad (8)$$

The result is the estimate $\widehat{Y'} = 0.61$ ‰, which is about 1/3 of the Bayes-Supplement 1 estimate. As already noted in [11], where Supplement 1 adds ($\hat{Y} = E[P]^2 + var[P]$), the frequentist approach subtracts ($\widehat{Y'} = \hat{P}^2 - var[\hat{P}]$).

Fig. 6 compares the risk function $R(P) = E[L(P)]$, for the MSE loss function $L(P) = (\hat{Y} - Y)^2 = [\widehat{P^2} - P^2]^2$, considering:

- the frequentist estimate (8)
- the frequentist estimate without correction of the bias, $\hat{Y} = P^2$
- the Bayesian estimate (7).

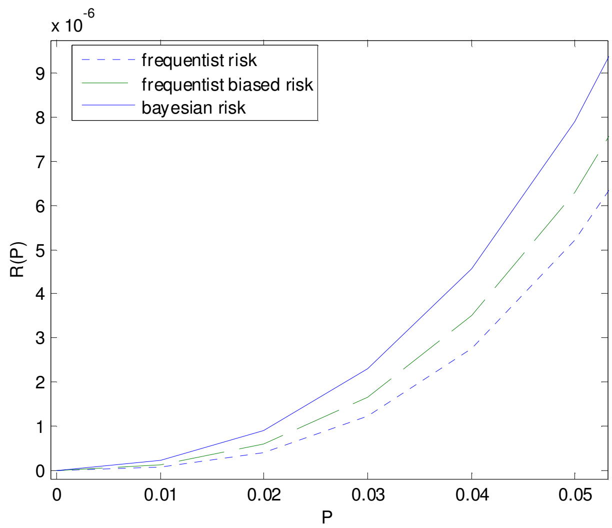

Fig. 6 – Risk function for the Bayesian and the frequentist estimates on the value of $Y = P^2$.

The figure shows clearly that the minimum risk (expected squared error) is associated with the frequentist estimate $\hat{Y} = 0.61$ ‰. The uncorrected frequentist estimate $\hat{Y} = 0.9$ ‰ is worse, and the Bayesian estimate $\hat{Y} = 1.53$ ‰ is a clear overestimation.

It must be highlighted that, *on average*, the Bayesian estimate is superior to the frequentist one, in a precise sense. If one has a set of welding machines with different values of P, such that P follows the distribution (4), than, on the long run, the Bayesian estimate will have an inferior MSE. But this is not a realistic situation and, as a matter of fact, if one observes 3 defective welds out of 100, the probability of having two consecutive defective welds is unlikely to be 1.53 ‰, and is likely to be, instead, 0.61 ‰.

C. How to propagate distributions in the objective Bayesian framework

The findings above does not mean that objective Bayesianism is incorrect. The problem arises from the absence of prior information, or, to say it more precisely, when *prior information* (available before taking the measurements) *has a negligible role with respect to that coming from the measurements*.

In this very common case, the Bayesian approach requires a "non-informative" prior, one that "let the data speak". Unfortunately, the prior that is good for the Bayesian estimate of *X*, overwhelms the data when estimating X^2 (or other nonlinear functions of *X*). In [7, p. 12], the procedure to obtain a correct Bayesian posterior distribution for squared quantities is described, and it implies the multiplication of the posterior pdf by the term $1/\sqrt{x}$.

Of course, the propagation of distributions described in the GUM – Supplement 1 is simpler and do not involve the subtleties of obiective Bayesianism. Many researchers have, in fact, observed that applying the propagation of distribution of Supplement 1 may lead to strange results. The solution is implementing, after the propagation, an additional numerical procedure, which is analogous to apply Bernardo's reference prior [12].

V. CONCLUSIONS

By examining the characteristics of the two rival approaches to statistical inference, the frequentist and the Bayesian ones, it is clearly understood why the latter has been preferred for setting the GUM, but it is also clear why this decision has caused problems. The central issue GUM deals with is the *propagation*, which is really the most difficult task in the Bayesian approach. It requires the definition of a new prior distribution for each measurement model, and an easy and automatic way to do it does not exist. In the frequentist scheme, instead, although convoluted and primitive, the propagation is extremely simple.

In light of the above observations, a doubt arises. Is a complex and delicate theory, as the objective Bayesianism, really suitable to define automatic procedures for the evaluation of estimates and uncertainties, and in particular for their propagation? Besides, how the uncertainty of frequentist estimates must be assessed? Should they be prohibited?

ACKNOWLEDGMENTS

This work has been partially funded by Project PON R&C 2007-2013 "RES NOVAE".

REFERENCES

[1] B. Efron, "Bayesians, Frequentists, and Scientists," *J. Am. Stat. Assoc.*, vol. 100, pp. 1–5, 2005.

[2] N. Giaquinto, "Incertezza di misura: teoria coerente o edificio da ricostruire? Parte 2: due approcci rivali alla teoria dell'incertezza," *Tutto_Misure*, vol. XV, no. 4, pp. 271–276, Dec. 2013.

[3] M. Bayes and M. Price, "An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.," *Philos. Trans.*, vol. 53, pp. 370–418, Jan. 1763.

[4] Clopper, C. J. and E. S. Pearson, "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, vol. 26, no. 4, pp. 404–413, Dec. 1934.

[5] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. New York: Springer, 1993.

[6] H. Jeffreys, "An Invariant Form for the Prior Probability in Estimation Problems," *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.*, vol. 186, no. 1007, pp. 453–461, Sep. 1946.

[7] J. M. Bernardo, "Reference Posterior Distributions for Bayesian Inference," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 41, no. 2, pp. 113–147, 1979.

[8] A. R. Syversveen, "Noninformative Bayesian Priors. Interpretation And Problems With Construction And Applications.," 1998.

[9] C. Stein, "An Example of Wide Discrepancy Between Fiducial and Confidence Intervals," *Ann. Math. Stat.*, vol. 30, no. 4, pp. 877–880, 1959.

[10] J. M. Bernardo, "Noninformative priors do not exist: A discussion.," *J Stat. Plan. Inference*, vol. 65, pp. 159–189, 1997.

[11] F. Attivissimo, N. Giaquinto, and M. Savino, "A Bayesian Paradox and its Impact on the GUM Approach to Uncertainty," *Measurement*, 2012.

[12] A. B. Forbes, "An MCMC algorithm based on GUM Supplement 1 for uncertainty evaluation," *Measurement*, vol. 45, no. 5, pp. 1188–1199, Jun. 2012.