USING SOME "PRIOR" WHEN THERE IS NO PRIOR

George Kahrimanis
anakreon@hol.gr
Southern Univeristy, Baton Rouge, Louisiana, USA

Abstract

If a prior probability does not exist or is controversial, can we formulate probabilistic inference based only on the properties of the measurement? In the case of plain 'location' measurements the answer to this is known. Here we extend that answer to address the general problem of estimating a scalar parameter from a small data sample. That measurement is treated as if resulting from perturbations of a location measurement. A proof is provided, that always a "posterior" without a prior exists and is equal to the posterior based on "Jeffreys' prior". The latter density is utilised in this way only, without assuming that it is the real prior, because Jeffreys' prior may be updated when we combine measurements, or when we revise the stopping rule. Prior-free inference is meaningful only if no prior exists or if it is provisionally suppressed. We discuss in what kind of problems this method is applicable.

1 INTRODUCTION

Where there is a prior probability [. . .] you don't have to be a Bayesian [to apply it]: that's because the idea of using a prior probability makes sense because there is *a prior probability. The problem is, do you want to put in a prior probability for something like the mass of the Higgs where there is no prior probability?* Fred James, [1].

A review of related literature is in [2]. For an example of obtaining a posterior-like *pdf* without a prior, consider a measurement of the angular distance of two stars in the sky, picked at random: if we know the precision of the measurement, then this is also our uncertainty about the angle. This is a coarse statement of an experimental scientist's intuitive interpretation of a plain 'location' measurement, as long as no prior beliefs are sustained about the location parameter. Related to this interpretation, certain statistical foundations have been discussed (fiducial, structural, pivotal)[1] but usually in such terms that correspondence to a physicist's intuition is not clear. Here we shall apply only Bayesian notions; not by forcing a prior on every analysis, but by considering the implications of lacking a prior.

It may happen, as a coincidence, that the prior-free "posterior" is numerically equal to the posterior that would be obtained if a certain prior were assumed. This is true in the case of location measurements (reviewed in the next section), for which the prior with the matching posterior is uniform in the location parameter. It would be convenient if for any problem the prior-free posterior could be calculated by multiplying the likelihood by some "prior with matching posterior". That is, we would like to prove that (a) the prior-free posterior exists always, and (b) it is equal to the product of the likelihood by a certain density, 'the prior with the matching posterior'. These conjectures are nontrivial, because we cannot invoke Bayes' Theorem directly: that theorem presupposes believing a certain prior probability, not merely believing that it is a mathematical object which leads to pragmatically adequate inference.[2]

The aim of this article is to justify the conjectures and find the formula for the general case. The main idea is to consider the generic problem of interest as if it were the result of succesive perturbations, starting with some location measurement. A simple example is the measurement of the angular distance of two stars in the presence of faint clouds. Even a counting experiment can be so linked to some

[1]About those, see [3] and cited references, esp. Barnard, Fisher, and Fraser.

[2]The phrase 'uninformative prior' is an oxymoron [5] if it is adopted uncritically.

location measurement, imagining that, by means of a sequence of perturbations, continuous probability is gradually deformed to approach a sum of weighted delta functions.

We shall also discuss in brief the question in what circumstances prior-free posteriors apply. The problem of vector parameter estimation is not addressed in this work, to postpone dealing with questions related to hierarchic grouping of parameters [2].

2 THE RELEVANCE OF JEFFREYS' PRIOR FOR SMALL DATA SAMPLES

2.1 Survey of known results

Definition. If a measurement provides a single datum x for estimating a scalar parameter θ, the term 'location parameter' means that, in this parametrization, the measurement error $e \equiv x - \theta$ is an additive noise of known *pdf* $\mathrm{Pr}_E(e|b) = f(e)$ (where 'b' stands for the evidence that establishes $f(e)$). E is regarded as causally unrelated to anything else in this model.

When we run the measurement, or a simulation, for calibration purposes, θ is known; in an actual measurement it is considered fixed but unknown, corresponding to a Bayesian random variable Θ. If θ is assumed fixed, the possible values of x correspond to a random variable X of probability density

$$\mathrm{Pr}_X(x|\theta, b) = \mathrm{Pr}_E(x - \theta|\theta, b) = f(x - \theta). \tag{1}$$

(Densities are represented by functions in the stated parametrization.)

After an actual measurement, *if* along with the datum x we also possess some information H leading to a prior $\pi(\theta) = \mathrm{Pr}_\Theta(\theta|H)$, Bayes' Theorem will lead to the posterior *pdf* for Θ: $\mathrm{Pr}_\Theta(\theta|x, H, b) \propto \pi(\theta)f(x-\theta)$. Note that the related *pdf* for the error E will be $\mathrm{Pr}_E(e|x, H, b) = \mathrm{Pr}_\Theta(x - e|x, H, b) \propto \pi(x-e)f(e)$. The latter *pdf* can be viewed either as the result of change of variable or as the posterior of an alternate Bayesian estimation: to an instrument maker, the quantity of interest is e. He has established prior beliefs $\mathrm{Pr}_E(e|b) = f(e)$ about it and is ready to update those beliefs, when he obtains the datum x and the prior $\pi(\theta)$ for Θ. Yet if no prior is available for θ, the instrument maker will remain with his prior for the error, because there is no intermediate step in Bayesian inference. Taking the datum x into account, his prior beliefs about the error are turned into beliefs about the measured parameter, by means of changing variable to $\Theta \equiv X - E$:

$$\mathrm{Pr}_\Theta(\theta|x, b) = \mathrm{Pr}_E(x - \theta|x, b) = f(x - \theta). \tag{2}$$

In plain words, if the precision of a measurement is known, the uncertainty of the outcome is its mirror image, as long as no prior for Θ has yet been taken into consideration.[3] This is a provisional result (not a marginal *pdf*) whether in a Bayesian or in a frequentist approach, *assuming a location measurement*. If a prior $\pi(\theta)$ is admitted, this *pdf* is superseded by the posterior.

This result is equal to the fiducial *pdf* in this case. It also agrees with the posterior that would be obtained if one assumed a prior uniform in the location parameter (we have not done so). In this article the latter density will be called the *false prior* of the problem, to emphasise that (like a false ceiling or false teeth) it has a function but is not the real thing.

In any specific problem, existence of a false prior simply means that we can achieve prior-free probabilistic inference by means of multiplying the likelihood by the false prior (as it happens; not because Bayes' Theorem applies). Non-existence of a false prior would imply either that no prior-free "posterior" (*e. g.* like that of Eq. 2) exists, or a situation in which the ratio of that *pdf* to the likelihood would depend not only on θ but also on the data.

"Virtually all default Bayesian methods recommend this conditional prior, as do various "structural" and even frequentist approaches" ([4], Sec. 4.3 in the preprint, referring to a more general case.)

[3]The same change of variable that is used to derive Eq. 2 has also been noted to relate the two prosteriors mentioned above. An equivalent change of variable has been used to derive Eq. 1, which is the mirror image of Eq. 2. This feature is masked if the definition of location measurement is merely a formal assumption of the two ends of Eq. 1, as is usual.

Therefore, in the case of a location measurement at least, it is legitimate for a physicist to follow plain intuition and consider the option of no prior. Of course we must examine when this option is appropriate; conversely, in what circumstances 'educated guessing' can be justified. These questions are left for a later section. Here we shall concentrate on generalising the method beyond the case of a plain location measurement.

Rather than a single real number x, the data may be an array $\mathbf{x}$ of n real numbers. Provided that the related error array is regarded as independent additive noise of known n-dimensional distribution, the same arguments apply. That is, prior-free probabilistic inference can be obtained, and the related false prior is again uniform in θ.

A special case of array location measurement is obtained from the combination of a set of measurements of the same location parameter. Because a false prior is not a real prior, we cannot (yet) conclude from the existence of a false prior for a single location measurement that the same (or any) false prior would apply to a combination of such measurements. But this conclusion is obtained directly, because a combination of location measurements is a special case of an array location measurement.

The unison of different statistical foundations in the case of location measurements does not extend to the general problem. Only with asymptotic approximations, assuming great accumulation of data, we find an agreement again. This is due to the existence of a reparametrization of θ, for which the condition of location measurement is asymptotically satisfied in some sense, as several authors have observed (though not always in clear terms, as noted in [2]).[4] Jeffreys' prior is defined as a density that is uniform in that parametrization.[5] In the original parametrization, it is (using $\Pr(\mathbf{x}|\theta)$ as shorthand for $\Pr_{\mathbf{X}}(\mathbf{x}|\theta, \text{model})$)

$$\pi_J(\theta) \propto \left[\int d^n x \; [\partial \Pr(\mathbf{x}|\theta)/\partial\theta]^2 \,/\, \Pr(\mathbf{x}|\theta)\right]^{0.5} = \left[-\int d^n x \; \Pr(\mathbf{x}|\theta) \left[\partial^2 \ln(\Pr(\mathbf{x}|\theta))/\partial\theta^2\right]\right]^{0.5} . \tag{3}$$

The same density has also been defended in terms of its utility, using decision-theoretical arguments with asymptotic assumptions [2]. It has some remarkable properties. It is invariant under reparametrizations of θ. Also, if we consider a combination of two measurements, Jeffreys' prior is updated, using the combined $\Pr(\mathbf{x}|\theta)$. Note that a real prior would not be updatable; the probability would be updated but the prior would remain fixed, unless the prior information were revised. Yet if we combine two measurements of the same kind (that is, with identical statistical models $\Pr(\mathbf{x}|\theta)$) the updating of Jeffreys' prior leaves it unchanged.[6]

Jeffreys' prior is meant for the asymptotic case, but in the rest of this work we shall extend the range of its application. As a starting point, note that for location measurements the false prior coincides with Jeffreys' prior.

[4] Here is a condensed account. As data accumulate we expect that the sample distribution almost certainly approaches the theoretical $\Pr(\mathbf{x}|\theta)$, therefore we think of the data sample as a lump that can be represented by means of a single number, if we disregard the asymptotically negligible statistical noise. That number may be the maximum-likelihood estimate generated by that sample. For any fixed θ_0, the distribution of the m-l estimate approaches the Gaussian shape, with variance approaching zero asymptotically. In general that variance also depends on θ_0, but if we use an appropriate reparametrization ('variance-stabilizing') of Θ the variance of the m-l estimate can become independent of the assumed true value, regardless of the sample size, provided that it is large enough. Because of asymptotic uniformity in the shape of the distribution of the m-l estimate about any tentative value θ, we have approximately formal equivalence with the definition of a location parameter.

[5] One way of deriving the form of a variance-stabilizing parametrization: consider grouping the accumulated data into bins of $\mathbf{x}$, for which the Gaussian approximation is applied, and find a condition required for the constancy the second derivative of the log-likelihood.

[6] To see why this is so, note that the asymptotic case, which consists of a large number of repetitions of the same measurement, can also be thought of as repeating *pairs* of the same measurement – or triads, and so on.

2.2 Prior-free inference in the general case

We have seen cases (combinations of location measurements, and the asymptotic case) in which the result can be obtained by multiplying the likelihood by a false prior, but we have not yet proved that this method is applicable also in the general case. As stated in the Introduction, to study the general, small-sample case we shall regard any measurement as the product of a sequence of small perturbations, starting either with some location measurement or with a finite combination of location measurements (depending on whether our measurement provides a single datum or an array of data).

In this section we shall consider a proof that, in the limit of infinitely many successive infinitesimal perturbations, the property "prior-free inference being equivalent to using Jeffreys' prior" is preserved to the first order in the sum of the strength of perturbations. But in this limit all higher-order effects vanish; the first-order theorem is all we need to prove.

Here we do not scrutinise the assumption that any problem of interest can be derived by means of perturbations starting with location measurements; let us just suppose so. We also need to assume that the *existence* of a prior-free result is not canceled by a perturbation. For a physicist this assumption is self-evident: it would be bizarre if prior-free inference could be extracted from some measurement but were precluded for an infinitesimally perturbed measurement.

Statistical analysis of all kinds is based on the following assumption. If two measurements are described by equivalent statistical models, *i.e.*

$$(\forall\,(\mathbf{x},\theta))\;\;\mathrm{Pr}_{\mathbf{X}_A}(\mathbf{x}|\Theta_A=\theta,\mathrm{Model}\,A)\;=\;\mathrm{Pr}_{\mathbf{X}_B}(\mathbf{x}|\Theta_B=\theta,\mathrm{Model}\,B), \tag{4}$$

and we presume equivalent prior information (or no prior, as far as we are concerned) then the same data leads to the same inference in either case. For studying perturbations, we need to adapt this assumption for cases in which the equality is not exact but approximate, whether to a certain order in a power expansion, or in a probabilistic sense. We shall regard Model B as a variation of Model A, and will be concerned with determining the corresponding variation of the prior-free posterior.

Proof. If we supposed the existence of a transformation (that is, reparametrization) of the data space along with a transformation of the parameter space, such that the transformed Model B be equivalent to the unaltered Model A, then the prior-free posterior for Model B would be defined in the new parametrization just by copying the corresponding density of Model A. Yet in general such an exact matching does not exist.

We can define "approximately matching" transformations if we take into account the prior-free posterior in Model A,[7] $\mathrm{Pr}_\Theta(\theta|\mathbf{x},\mathrm{Model}\,A)$ and use it to determine average difference of the two sides in Eq. 4, as well as the variance of the distribution of that difference: for "approximate matching", the average is required to be zero and the variance not far from the minimal possible, that is, within a fixed fractional tolerance, say 80%. This definition depends on the degree of tolerance (but this will be seen to be of no consequence). Note that, even if a best matching were defined uniquely by the requirement of minimal variance, any other transformation that results to a variance within 1% from the minimal one would present nearly equal claim to approximately matching Model A.

With any aproximate matching, if we copy the prior-free posterior $\mathrm{Pr}_\Theta(\theta|\mathbf{x},\mathrm{Model}\,A)$ into a density of the parameter of the transformed Model B, as if the two models were matched exactly by way of this transformation, we will obtain the prior-free posterior not of Model B but of a mutated model. (This observation will be useful when we consider varying models, so that, in the limit, the minimal variance approaches zero.)

Now let us think of Model B as the result of a perturbation of Model A, of strength t. The perturbed model can be expressed as ${}^t\mathrm{Pr}_{\mathbf{X}}(\mathbf{x}|\Theta=\theta)={}^0\mathrm{Pr}_{\mathbf{X}}(\mathbf{x}|\Theta=\theta)+t\,g(\mathbf{x};\theta)+O(t^2)$, where t is the perturbation strength, and g is some density with regard to $\mathbf{x}$ and also a function of θ. In this section, the sign "$\approx$" denotes an expansion to the first order.

[7] In studying perturbations, one makes use of an approximate result in order to calculate the next correction.

We take advantage of tranformations of data array $\mathbf{x}$ and of parameter θ, to vary the form of ${}^t\mathrm{Pr}_{\dots}(\mathbf{x}|\theta)$, trying to match the unperturbed form ${}^0\mathrm{Pr}_{\dots}(\mathbf{x}|\theta)$. Any transformation from $(\mathbf{x},\theta)$ to $(\mathbf{y},\phi)$ satisfies relations like $\mathbf{y} = \mathbf{x} + t\,\mathbf{e}(\mathbf{x})$ and $\phi = \theta + t\,f(\theta)$, for some scalar function f and array function $\mathbf{e}$, which are at our disposal to define, provided that the resulting transformations are one-to-one.

When the data array is transformed, probability density is multiplied by the Jacobian determinant, J. On the basis of the expansion $|\mathbf{I} + \epsilon\mathbf{A}| = 1 + \epsilon\mathrm{Tr}(\mathbf{A}) + O(\epsilon^2)$, we obtain $J \approx 1 - t\sum \partial e_i/\partial x_i$.

The transformed probability density is ${}^t\mathrm{Pr}_{\mathbf{Y}}(\mathbf{y}|\Phi = \phi) \approx J\ {}^t\mathrm{Pr}_{\mathbf{X}}(\mathbf{y} - t\mathbf{e}(\mathbf{x})|\Theta = \phi - tf(\phi))$ $\approx {}^0\mathrm{Pr}_{\mathbf{X}}(\mathbf{y}|\Theta = \phi) + t\,[g(\mathbf{x};\phi) - {}^0\mathrm{Pr}_{\mathbf{X}}(\mathbf{y}|\Theta = \phi)\sum\ \partial e_i(\mathbf{y})/\partial y_i - \sum\ \partial\,{}^0\mathrm{Pr}_{\mathbf{X}}(\mathbf{y}|\Theta = \phi)/\partial y_i\ e_i(\mathbf{y})$ $-\ \partial\,{}^0\mathrm{Pr}_{\mathbf{X}}(\mathbf{y}|\Theta = \phi)/\partial\phi\ f(\phi)]$.

According to the earlier definition, the pair of transformations specified by f and $\mathbf{e}$ constitute an "approximate matching" to the first order iff the sum in square brackets is of zero average and of nearly (within some fixed fractional tolerance) minimal variance, with regard to the prior-free posterior of the model with t=0. With any such pair of f and $\mathbf{e}$, let us copy the prior-free posterior of the unperturbed model into a density of ϕ, as if the matching were exact; the result can be thought of as the prior-free posterior of a mutated model, which is defined by dropping the t term in the last equation above. Here we have assumed that the assumption expressed in Eq. 4 also holds as a first-order expansion.

If we happen to know that Jeffreys' prior provides prior-free inference for t=0, this must also be true to the first order in t for mutated models. Note that Jeffreys' prior is invariant under transformations of the parameter space as well as of the data space.

The difference of any mutated model from the non-mutated one (the latter is the result of perturbation with strength t) is zero in the zero-order expansion, zero-in-the-average in the first-order expansion in t, and with variance of the second order. In general for each t there is an infinite number of such mutated models, each corresponding to an approximately matching transformation; note that the difference between any two such models has the same expansion properties as the difference from the non-mutated model.

We return to studying the sequence of perturbations which leads from a location model to the real model. This correspondence becomes exact in an appropriate limit, as the number of perturbations increases without bound while the strength of each perturbation goes to zero. Consider the mutated models that arise in the first perturbation; each of those mutated models leads to a family of mutated models in the second perturbation, and so on.

In the final perturbation, an extended family of mutated models is assembled; the difference of any two such models is zero in the zero-order, zero-in-the-average in the first order of the sum of perturbation strengths, and of variance that vanishes in the limit, because it is $O(t^2)$. All higher-order terms vanish identically in this limit. The same can be said for the difference between any mutated model and the non-mutated model. Therefore, in the limit, the whole final extended family of mutated models shrinks to approximate the non-mutated model. This result does not depend on the tolerance we set in the definition of approximate matching: with wider tolerance we would get larger families, but the limit would be the same.

According to the above considerations, for each mutated model, the prior-free posterior is already defined, to the first order in the sum of perturbation strengths. But higher-order terms vanish in the limit, therefore the definition becomes exact in that limit. Because of the shrinking of the final extended family of mutated models in the limit, this is the definition of the prior-free posterior for the real model.

As noted in the previous section, the false prior of any location model coincides with Jeffreys' prior. According to the above proof the same holds for each generation of mutated models in the sequence of perturbations (that is, to the first order with vanishing higher-order effect) therefore it also holds for the real model.

2.3 Application to problems in high-energy physics

In HEP, where data samples are often small or empty, such considerations are relevant. In the case of a counting experiment of a pre-set duration, Jeffreys' prior for the Poissor rate μ is $\propto \mu^{-0.5}$. If a search has returned an empty sample, an upper limit based on this prior is about half of the corresponding upper limit that is based on a prior uniform in μ. If the sample size is n, the Bayesian average for μ is $n + 0.5$ rather than $n + 1$. More important than implementing these adjustments, we tackle the 'anxiety' whether using a prior makes sense at all.

3 WHEN DO WE NEED PRIOR-FREE INFERENCE

Here is an example. In a geodesic survey involving three mountain peaks, we are interested in the sum, θ, of the angles of that triangle. Besides the significance of the measurement for testing space-time theories, suppose that we are also interested in estimating θ, even if only for the sake of quoting a result.

Every gravitational theory of space-time provides a model. If someone believes that Euclidean geometry is necessarily true, then his prior is a delta function, $\delta(\theta - \pi)$. Strictly speaking, he cannot even define likelihood for $\theta \neq \pi$ because such a definition would be based on a contradiction.

In the approach based on Einstein's General Relativity, θ is expressed in terms of related magnitudes M_i (*e.g.* the mass of the Earth) and physical constants C_j. The pre-data information on these parameters is expressed as a joint probability density $\pi(\{M_i\}, \{C_j\})$, from which we can calculate the pre-data probability density of θ. The measurement updates the former probability, therefore also updates, indirectly, the *pdf* regarding θ. If the measurement is not precise enough, the result will be dominated by pre-data information. Note that, in the exact treatment, θ is not a parameter of the main calculations: we do *not* define likelihood of θ using probability $\mathrm{Pr}_X(x|\theta, \text{model})$. The same considerations apply to any other space-time gravitational theory.[8]

If we cannot even define a model-independent likelihood $\mathrm{Pr}_X(x|\theta, *)$ for all implicated models, there is no way to achieve universally valid estimation of θ, even if we had agreed on some compromise for prior. (At most we can perform a significance test for each model separately.) However, from the experimentalists' point of view there *is* an empirically derived probability $\mathrm{Pr}_X(x|\theta, \text{apparatus})$, based on calibrations and trials of the three angle measuring devices inside a small laboratory – provided that we disregard the physical meaning, in each theory, of the parameter measured by means of the survey.

From the point of view of each theory, the above probability is only an approximation. Its validity is explained by the applicability of Euclidean geometry in the small range of the laboratory, because the space curvature deduced from the approximately known M_i and C_j is not too large. Yet, if the experimentalist's formula is considered for values of θ quite removed from π, space curvature would be large enough to have an effect even in lab calibration runs, according to the theory.

From the experimentalist's point of view this concern (if he is aware of it) is purely theoretical because there is ample empirical evidence that Euclidean geometry is in agreement with measurements inside the laboratory. After all, if the survey returned such a value for θ, the model would be discredited, for failing to explain the applicability of Euclidean geometry inside the lab.

Now the problem can be formulated simply as estimating the experimentalist's θ from datum x. Note that the agreed form of $\mathrm{Pr}_X(x|\theta, \text{apparatus})$ presupposes disregarding the physical meaning of θ. Accordingly, there is no prior for the experimentalist's θ. (Suppose for example that a general relativist adopted his pre-data *pdf* for θ as a prior; aside from the logical inconsistency of mixing the empirical likelihood with a theoretical prior, note that this would be an inaccurate treatment, because it would ignore correlations in $\pi(\{M_i\}, \{C_j\})$.) In other words, the prior-free option is the only way to formulate

[8] It is possible to invent additional parameters and contrive some general model that encompasses as special cases General Relativity and Euclidean geometry. Unless this general model is based on some plausible guiding principle, the definition of likelihood in the intermediate cases will be deemed arbitrary. That is, an artificial general theory cannot subsume the particular theories in this discussion.

parameter estimation that does not depend on the assumptions of a particular model.

This is not to say that there is no place for any private or tentative assumption. A professional practitioner may apply educated guessing in making qualified decisions, diagnoses, and forecasts. A civil engineer usually assumes without testing that Euclidean geometry is an adequate approximation. For another example, consider a physician who wants to interpret the results of the HIV test of a patient who has been classified as a regular morphine user. Because the reliability of the result is limited, the physician wants to take into account the rate of HIV infection in that group, but he does not know that rate. Yet he knows the rate of infection among heroin users, and he believes that the mechanism that accounts for the high rates of infection in that group also applies in the case of morphine users. He calculates his posterior probability of HIV infection for this patient using the assumption that the two group rates are approximately equal.

But when we estimate a parameter we must revert to assumptions that are not disputed in the context of this measurement.[9] In the context of the last example, after we sample a group of people to estimate the group rate of HIV infection, when we analyse the result we may need to disregard our prior beliefs about the mechanism of infection, if they are not precise enough, or not shared by every practitioner.

4 CONCLUSIONS

In the case of plain location measurements, intuitive formation of prior-free probabilistic inference is justified from a Bayesian point of view, as long as no prior is admitted. Applying successive perturbations, we can extend the notion of prior-free result to the general case of scalar parameter estimation.

The calculation of any prior-free result can be obtained by multiplying the likelihood by Jeffreys' prior. The latter is not the real prior of the problem, only a calculational tool. That is, we do not apply Bayes' Theorem; we only take advantage of a shortcut. When we combine two different measurements we must revise the prior, to be Jeffreys' prior of the combined statistical model. The stopping condition of the sampling affects Jeffreys' prior, in contrast to any case in which a real prior is applied.

Qualified conjectures, even if suitable for professional decision making, may be questionable in the context of parameter estimation. For such cases, and when there is no prior available, the no-prior option is needed. This approach requires that $\mathrm{Pr}_X(x|\theta, \text{apparatus})$ can be derived within some adequate approximation (that is, in comparison to the minimal standard deviation) on the basis of presumably undisputed information and assumptions, like calibration data.

ACKNOWLEDGEMENTS

The presentation of this paper was partly funded by the Onassis Foundation, at the recommendation of Mrs. Alcestis Soulogiannis, of the Ministry of Culture of Greece. I would like to thank Günter Zech for stimulating exchanges which have motivated this work.

References

[1] P. Clifford, presentation and discussion, in the Workshop on Confidence Limits at CERN (2000), in CERN Yellow Report 2000-005; posted at <http://documents.cern.ch/archive/electronic/cern/preprint/open/open-2000-207.pdf>.

[2] R. E. Kass and L. Wasserman, J. Amer. Statist. Assoc. **91** (1996) 1343-1370; related Carnegie Melon Report <http://www.stat.cmu.edu/www/cmu-stats/tr/tr583/tr583.ps>.

[9] A prior belief can be the basis for doubting the validity of an experiment, but if we ever come to admit the fairness of that experiment then an objective formulation of the result would use only undisputed asumptions.

[3] D. Salomé, Ph.D. Thesis (1998) Rijksuniversiteit Groningen; <http://www.ub.rug.nl/eldoc/dis/science/d.salome/>.

[4] J. Berger, B. Liseo, and R. Wolpert, Statistical Science **14** (1999) 1-28; preprint <http://www.isds.duke.edu/~berger/papers/brunero.ps>.

[5] G. L. Wallstrom, Ph.D. Thesis (1999) University of Minnesota; <http://www.stat.umn.edu/PAPERS/theses/garrick_thesis.ps>.