Rev.R.Acad.Cienc.Exact.Fis.Nat. (Esp)
Vol. 93, N.º 3, pp 311-319, 1999
Monográfico: Métodos Bayesianos en las Ciencias

OVERCOMING PRIORS ANXIETY

(Subjective Bayesian theory/subjective priors/high energy physics/measurement of uncertainty)

GIULIO D'AGOSTINI*

* Dipartimento di Fisica dell'Università «La Sapienza» and Instituto Nazionale di Fisica Nucleare (INFN) P. le Aldo Moro 2, I-00185 Roma (Italy). dagostini@roma1.infn.it.URL:www-zeus.roma1.infn.it/~agostini/

ABSTRACT

The choice of priors may become an insoluble problem if priors and Bayes' rule are not seen and accepted in the framework of subjectivism. Therefore, the meaning and the role of subjectivity in science is considered and defended from the pragmatic point of view of an «experienced scientist». The case for the use of subjective priors is then supported and some recommendations for routine and frontier measurement applications are given. The issue of reference priors is also considered from the practical point of view and in the general context of «Bayesian dogmatism».

RESUMEN

Superando la ansiedad producida por las distribuciones iniciales

La elección de distribuciones iniciales puede resultar un problema insoluble si las distribuciones iniciales y la regla de Bayes no son aceptadas en un marco subjetivista. Consecuentemente, consideramos el significado y el papel de la subjetividad en las ciencias desde el punto de vista de pragmático de un «científico experimentado». Se defiende el uso de distribuciones iniciales subjetivas, y se ofrecen recomendaciones para aplicaciones, tanto rutinarias como punteras. El tema de las distribuciones de referencia es asimismo considerado, tanto desde el punto de vista práctico como en el contexto general del «dogmatismo Bayesiano».

1. INTRODUCTION

The main resistance scientists have to Bayesian theory seems to be due to their reaction in the face of words such as «subjective»,«belief» and «priors» (to which the word «bet» might also be added). These words sound blasphemous to those who pursue the ideal of an objective Science. Given this premise, it is not surprising that frequentistic ideas, which advertise objective methods at low cost in a kind of demagogical way, became popular very quickly and are still the most widely used in all fields of application, despite the fact that they are indefensible from a rational point of view. As in commercials, what often matters is just the slogan, not the quality of the product, at least in the short term. And advertised objective methods are certainly easier to sell than subjective ones. When one adds to these psychological effects yet others based upon political reasons [see, for example, the very interesting philosophical and historical introduction to Lad, 1996], life gets really hard for subjective probability. Moving from the slogan to the product, it is not difficult to see that, if they were to be taken literally, frequentistic ideas would lead nowhere. Indeed their success seems due to a mismatch between what they state and how scientists interpret them in good faith. In other words, frequentistic methods make sense only if they are —when they can be— reinterpreted from a subjective point of view. Otherwise they may cause serious mistakes to be made. In recent years I have investigated this question among particle physicists (D'Agostini, 1998, 1999). For the convenience of the reader, I report here the main conclusions contained in D'Agostini (1998):

- there is a contradiction between a cultural background in statistics and the good sense of physicists; physicists' intuition is closer to the Bayesian approach than one might naïvely think;
- there are cases in which good sense alone is not enough and serious mistakes can be made; it is then that the philosophical and practical advantages offered by the Bayesian approach become of crucial importance;
- there is a chance that the Bayesian approach can become widely accepted, if it is presented in a way which is close to physicists' intuition and if it can solve the «existential» problem of reconciling two

aspects which seem irreconcilable: subjective probability and the honest ideal of objectivity that scientists have.

This last point was just sketched in the original paper, and I would like to discuss it here in a bit more detail, and to relate it to the «problem» of priors, the main subject of this article. I think, in fact, that it is impossible to talk about priors without putting them into the framework to which they belong. Only when one is aware of the role they have in Bayes' theorem, and of the role of Bayes' theorem itself, can one have a relaxed relationship with them. Once this is achieved, depending on the specific problem, one may choose the most suitable priors or ignore them if they are irrelevant; or one may decide, instead, that priors are so relevant that only Bayes' factors can be provided; alternatively one may even skip the Bayes' theorem altogether, or use it in a reverse mode to discover which kind of priors might give rise to the final beliefs that one unconsciously has. These situations will be illustrated by examples.

Before going any further, some clarifications are in order. First, my comments will be from the viewpoint of the «experienced scientist» (i.e. the scientist who is used to everyday confrontation with real data); this point of view is often neglected, since priors (and questions of subjectivity/objectivity) tend to be debated among mathematicians, statisticians and philosophers. Second, since I am an experimental particle physicist, I am aware that my knowledge about the literature concerning the arguments I am going to talk about is necessarily limited and fragmentary. I therefore apologize if people who may have expressed opinions similar to those stated in this paper are not acknowledged here.

2. SUBJECTIVE DEGREES OF BELIEF AND OBJETIVE SCIENCE

The question *«can subjective degrees of belief build an objective Science?»* is subtle. If we take it literally, the answer is NO. But this is not because of the subjective degrees of belief in themselves. It is simply because, from a logical point of view, «objective Science» is a contradiction in terms, if «Science» stand for Knowledge concerning Nature, and «objective» for something which has the same logical strenght as a mathematical theorem. This has been pointed out many times by philosophers, the strongest defence of this point of view being due to Hume (1748), to whom there is little to reply.

If, instead, «objective Science» stands for what scientists refer to by this expression, the question hecomes a tautology. In fact, using Galison' (1987), *«experiments begin and end a matrix of beliefs. .. beliefs in instrument types, in programs of experimental enquiry, in the trained, individual judgements about every local behaviour of pieces of apparatus...»*. Any scientist knows already that the only objective thing in science is the reading of digital scales. When we want to transform this information into scientific knowledge we have to make use of many implicit and explicit beliefs.

However, many scientists are reluctant to use the word «belief»[1] for professional purposes. It seems to me that the reason for this attitude is due to a misuse of the word «belief», which has somehow led to a deterioration of its meaning. In this connection I think a few remarks are of particular importance. The first is that we should have Hume's distinction between «belief» and «imagination» clear in mind (Hume, 1748). Then, once we agree on what «belief» is, and on the fact that it can have a degree, and that this degree depends necessarily on the subject who evaluates it, another important concept which enters the game is that of de Finetti (1974) «coherent bet». The «coherent bet» plays the crucial role of neatly separating «subjective» from «arbitrary». In fact, coherence has the normative role of forcing people to be honest and to make the best (i.e. the «most objective») assessments of their degree of belief[2]. Finally comes Bayes' (1764) rule, which is the logical tool for updating degrees of belief.

In my opinion there is a really good chance that this way of presenting the Bayesian theory will be accepted by scientists. In fact the ideal of objectivity is easily recovered, although in terms of *intersubjectivity*, if scientific knowledge is regarded as a very *solid Bayesian network* (Pearl, 1988) [Galison (1987) «matrix of beliefs»], based on centuries of experimentation, with *fuzzy borders* which correspond to the areas of current investigation.

3. CHOOSING PRIORS: FEAR, MISCONCEPTION AND GOOD FAITH

Once we have specified the exact meaning of each of the ingredients entering probabilistic induction (degree of belief —coherent bet— Bayes' rule), there should, in principle, no longer be a problem. However, all Bayesians know by experience that the most serious concerns scientists have are related to the choice of priors (sometimes due to real technical problems, but more often due only to «prejudices on priors»). In fact, practitioners can avoid talking about «degree of belief» in their papers, replacing it by the nobler term «probability»; they can accept the use of Bayes' theorem, because it is a theorem; but it seems they cannot escape from priors. And they often get stuck, or simply go back to «objec-

[1] But many other scientists, usually prominent ones, do. And, paradoxically, objective science is, for those who avoid the word «belief», nothing but the set of beliefs held by the most influential scientists in whom they believe...

[2] The coherence is also important to avoid making the confusion between «belief» and «convenience» (or «wish»). In other words, the tasks of assessing probability and of decision making shoud be kept separate.

tive» frequentistic methods. In fact, the choice of the prior is usually felt to be a vital problem by all those who approach the Bayesian methods with a purely utilitarian spirit, that is, without having assimilated the spirit of subjective probability. Some use «Bayesian formulae» simply because they «have been proved», by Monte Carlo simulation, to work in a particular application. Others seem convinced by the power of Bayesian reasoning, but they are embarrassed because of the apparent «arbitrariness» of the choice of priors.

It might seem that reference priors [see e.g. (Bernardo and Smith, 1994) and references therein, although in this paper I will refer only to Jeffreys' priors, the most common in Physics applications) have a chance of attracting people to Bayesian theory. In fact, reference priors enable practitioners to avoid the responsibility for choosing priors, and give them an *illusion of objectivity* analogous to that offered by frequentistic procedures (Berger and Berry, 1988). However I have some perplexity about uncritical use of reference priors, for philosophical, sociological and pactical reasons which I am now going to explain.

3.1. Bayesian dogmatism and its dangers

Although I agree, in principle, that a *«concept of a ''minimal informative'' prior specification - appropriately defined!»* (Bernardo and Smith, 1994) is valid, those who are not fully aware of the intentions and limits of reference analysis perceive the Bayesian approach to be dogmatic. Indeed, one can find indiscriminate use and uncritical recommendation of reference priors in books, lecture notes, articles and conference proceedings on Bayesian theory and applications. This gives to practitioners the impression that only those priors blessed by the official Bayesian literature are valid. This would be a minor problem if the use of reference priors, instead of more motivated ones, merely caused a greater or lesser difference in the numerical result. However, the question becomes more serious when the —perhaps unwanted— dogmatism is turned against the Bayesian theory itself. I would like to give an example of this kind wich concerns me very much, because it may influence the High Energy Physics community to wich I belong. In a paper wich appeared last year Feldman and Cousins (1998) stated that

> «For a parameter μ which is restricted to $[0, \infty]$, a common non-informative prior in the statistical literature is $p(\mu_t) = 1/\mu_t$ ... In contrast the PDG[3] description is equivalent to using a prior wich is uniform in μ_t. This prior has no basis that we know of in Bayesian theory».

This example should be taken really very seriously. The authors, in fact, use the pulpit of a prestigious journal to make it seem as if they understand deeply both the Bayesian approach and the frequentistic approach and, on this basis, they discourage the use of Bayesian methods (*«We then obtain confidence intervals which are never unphysical or empty. Thus they remove an original intention for the description of Bayesian intervals by the PDG»* (Feldman and Cousins, 1998).

So it seems to me that there is a risk that indiscriminate use of reference priors might harm the Bayesian theory in the long term, in a similar way to that which happened at the end o last century, as a consequence of the abuse of the uniform distribution. This worry is well expressed in Earman's conclusions to his «critical examination of Bayesian confirmation theory» (Earman, 1992):

> «We then seem to be faced with a dilemma. On the one hand, Bayesian considerations seem indispensable in formulating and evaluating scientific inference. But on the other hand, the use of the full Bayesian apparatus seems to commit the user to a form of dogmatism.»

3.2. Unstated motivations behind Jeffreys' priors?

Coming now to the specific case of Jeffreys' priors, I must admit that, from the most general (and abstract) point of view, it is not difficult to agree that *«in one-dimensional continuous regular problems, Jeffreys' prior is appropriate»* (Bernardo and Smith, 1994). Unfortunately, it is rarely the case that in practical situations the status of prior knowledge is equivalent to that expressed by the Jeffreys' priors, as I will discuss later. Reading «between the lines», it seems to me that the reasons for choosing these priors are essentially psychological and sociological. For instance, when utilized to infer μ (typically associated with the «true value») from «Gaussian small samples», the use of prior of the kind $f_0(\mu, \sigma) \propto 1/\sigma$ has two apparent benefits:

- first, the mathematical solution is simple (this reminds me of the story of the drunk under the streetlamp, looking for the key lost in the dark alley);
- second, one recovers the Student distribution, and for some it seems to be reassuring that a Bayesian result gets blessed by *«well established»* frequentistic methods. («We kow that this is the right solution», a convinced Bayesian once told me...)

But these arguments, never explicitly stated, cannot be accepted, for obvious reasons. I would like only to comment on the Student distribution. This is the «standard way» for handling small samples, although there is, in fact, no deep reason for aiming to get such a distribution for the posterior. This becomes clear to anyone who, having measured the size of this page twice and having

[3] PDG stands for «Particle Data Group», a committee that every second year publishes the *Review of Particle Properties* (14), a very influential collection of data, formulae and methods, including sections on Probability and Statistics.

found a difference of 0.3 mm between the measurements, then has to base his conclusion on that distribution. Any rational person will refuse to state that, in order to be 99.9 % confident in the result, the uncertainty interval should be 9.5 cm wide (any carpenter would laugh...). This might be the reason why, as far as I know, physicists dont's use the Student distribution.

Another typical application of the Jeffreys' prior is in the case of inference on the λ parameter of a Poisson distribution, having observed a certain number of events x. Many people have, in fact, a reluctance to accept, as an estimate to λ, a value which differs from the observed number of counts (for example, $E(\lambda) = x + 1$ starting from a uniform prior) and which is deemed to be distorted by the «distorted» frequentistic criteria used to analyse the problem [see e.g. D'Agostini, 1999]. In my opinion, in this case one should simply educate the practitioners about the difference between the concept of maximum belief and that of *prevision* (or expected value). An example in which the choice of priors becomes crucial, is the case where no counts are observed, a typical situation for frontier physics, where new and rare phenomena are constantly looked for. Any reasonable prior consistent with what I like to call the «positive attitude of the physicists who have pursued the research», allows reasonable upper limits compatible with the sensitivity of the experiment to be calculated (even a uniform priors is good for the purpose). Instead, a prior of the kind $f_0(\lambda) \propto 1/\lambda$ prevents use of probabilistic statements to summarize the outcome of the experiment, and the same result (0 ± 0) is obtained, independently of the size sensitivity and running time of the experiment.

I will return below to such critical situations which are typical of frontier science.

4. PRIORS FOR ROUTINE APPLICATIONS

Let us discuss now the reasons which indicate that experimentally motivated priors for «routine measurements» are quite different from Jeffreys' priors. This requires a brief reminder about how measurements are actually performed. I will also take the opportunity to introduce the International Organization for Standardization (ISO) recomendations concerning measurement uncertainty.

4.1. Unavoidable prior knowledge behind any measurement

To understand why an «experienced scientist» has difficulty in accepting a prior of the kind $f_0(\sigma) \propto 1/\sigma$ (or $f_0(\ln(\sigma)) = k$), one has to remember that the process of measurement is very complex (even in everyday situations, like measuring the size of the page *You* are reading *now*, just to avoid abstract problems):

- first one has to *define the measurad*, i.e. the quantity one is interested in;
- then one has to *choose the appropriate instrument*, one which has known properties, well-suited range and resolution, and in which one has some confidence, achieved on the basis of previous measurements;
- the *measurement* is performed and, if possible, repeated several times;
- then, if one judges that this is appropriate, one applies *corrections*, also based on previous experience with that kind of measurement, in order to take into account known (within uncertainty) systematic errors;
- finally[4] one gets a credibility interval for the quantity (usually a *best estimate* with a related *uncertainty*);

Each step involves some prior knowledge and, typically, each person who performs the measurement (be it a physicist, a biologist, or a carpenter) operates in his field of expertise. This means that he is well aware of the error he might make, and therefore of the uncertainty associated with the result. This is also true if only a single observation has been performed[5]: try to ask a carpenter how much he believes in his result, possibly helping him to quantify the uncertainty using the concept of the coherent bet.

There is also nother important aspect of the «single measurement». One should note that many measurements, which seem to be due to a single observation, consist, in fact, of several observations made within a short time: for example, measuring a length with a design ruler, one checks the alignment of the zero mark with the beginning of the segment to be measuremed several times; or, measuring a voltage with a voltmeter or a mass with a balance, one waits until the reading is well stabilized. Experts use unconsciously also information of this kind when they have to figure out the uncertainty they attribute to the result, althought they are unable to use it explicitly because this information cannot be accommodated in the standard way of evaluating uncertainty based on frequentistic methods (D'Agostini, 1998).

The fact that the evaluation of uncertainty does not necessarily come from repeated measurements has also

---

[4] This is not really the end of the story, if a researcher wishes his result to have some impact on the scientific community. Only if other people trust him will they use the result in further scientific reasoning, as if it were their own result. This is the reason why one has to undergo an apprenticeship during one's youth, when one must build up one's reputation (i.e. again beliefs) in the eyes of one's colleagues.

[5] This defence of the possibility of quoting an uncertainty from a single measurement has nothing to do with the mathematical games like those of (Rodríguez, 1999).

been recognized by the International Organization for Standardization (ISO) in its «*Guide to the expression of uncertainty in measurement*» (1993). There the uncertainty is classified.

«into two categories according to the way their numerical value is estimated:

A. those which are evaluated by statistical methods[6];

B. those which are evaluated by other means»;

Then, illustrating the ways to evaluate the «type B standard uncertainty», the *Guide* states that

«the associated estimated variance $u^2(x_i)$ or the standard uncertaintyu $u(x_i)$ is evaluated by scientific judgement based on all of the available information on the possible variability of X_i. The pool of information may include

— previous measurement data;

— experience with or general knowledge of the behaviour and properties of relevant materials and instruments;

— manufacturer's specifications;

— data provided in calibration and other certificates;

— uncertainties assigned to reference data taken from handbooks.»

It is easy to see that the above statements have sense only if the probability is interpreted as degree of belief, as explicitly recognized by the *Guide:*

«... Type B standard uncertainty is obtained from an assumed probability density function based on the degree of belief that an event will occur (often called subjective probability...).»

It is also interesting to read the concern of the *Guide* regarding the uncritical use of statistical methods and of abstract formulae:

«the evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement. The quality and utility of the uncertainty quoted for the result of a measurement therefore ultimately depend on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value».

This appears to me perfectly in line with the lesson of genuine subjectivism, accompanied by the normative rule of coherence.

4.2. Rough modelling of realistic priors

After these comments on measurement, it should be clear why a prior of the kind $f_0(\mu, \sigma) \propto 1/\sigma$ does not look natural. As far as σ is concerned, this prior would imply, in fact, that standard deviations ranging over many («infinite», in principle) orders of magnitude would be equally possible. This is unreasonable in most cases. For example, measuring the size of this page with a design ruler, no one would expect $\sigma \approx \mathcal{O}$ (10 cm) or $\approx \mathcal{O}$ (1 μm). As for μ, the choice $f_0(\mu) = k$ is acceptable until $\sigma \ll \mu$ (the so called Savage (1962) *principle of precise measurement*). But when the order of magnitude of σ is uncertain, the prior on μ should also be revised (for example, most of the directly measured quantities are positively defined).

Some priors which, in my experience, are closer to the typical prior knowledge of the person who makes *routine measurements* are those concerning the order of magnitude of σ, or the order of magnitude of the precision (quantifield by the variation coefficient $v = \sigma/|\mu|$). For example[7], one might expect a r.m.s. error of 1 mm, but values of 0.5 or 2.0 mm would not look surprising. Even 0.2 or 4 mm would look possible, but certainly not 1 μm or 10 cm. So, depending on whether one is uncertain on the absolute or the relative error, a distribution which seems suitable for a rough modelling of this kind of prior is a *lognormal* in either σ or v. For instance, the above example could be modeled with ln σ normally distributed with average 0 (= ln 1) and standard deviation 0.4. The 1, 2 and 3 standard deviation interval on σ/mm would be (0.7, 1.5), (0.5, 2.2) and (0.3, 3.3), respectively, in qualitative agreement with the prior knowledge.

In the case of more sophisticated measurements, in which the measurand is a positive defined quantity of unknown order of magnitude, a suitable prior would again be a normal (or, at limit, a constant) in ln μ (before the first measurement one may be uncertain on the order of magnitude that will be obtained), while σ is somehow correlated to μ (again v can be reasonably described by a lognormal). One might imagine of other possible measurements which give rise to other priors, but I find it very difficult to imagine a real situation for which Jeffreys' priors are appropriate.

[6] Here «statistical» should be seen as referring to «repeated observations on the same measurand», and not to general meaning of «probabilistic».

[7] For the sake of simplicity, let us stick to the case in which the fluctuations are larger than the instrumental resolution. Otherwise one needs to model the prior (and the likelihood) with a discrete distribution.

4.3. Mathematics versus good sense

The case of small samples seems to lead to an impasse. Either we have a simple and standard solution to a fictious problem, given by the Student distribution, or we have to face complicated calculations if we want to solve specifically the problem we have in mind, formulated by modeling experience motivated priors. I do not think that experimenters would be willing to calculte lognormal integrals to report the results of a couple of measurements. This could be done once, perhaps, to get a feeling of what is going on, or to solve an academic exercise, but certainly not as routine.

The suggestions sketched above were in the framework of the Bayes' theorem paradigm. But I don't want to give the impression that this is the only way to proceed. The most important teaching of subjective probability is that probability is always conditioned by a given status of information. The probability is updated in the light of any new information. But it is not always possible to describe the updating mechanism using the neat scheme of the Bayes' theorem. This is well known in many fields, and, in principle, there is no reason for considering the use of the Bayes theorem to be indispensable to assessing uncertainty in scientific measurements. The idea is to force the expert to declare (using the coherent bet) some quantiles in which he believes is contained the true value, on the basis of a few observations. It may be easier for him to estimate the uncertainty in this way, drawing on his past experience, rather than trying to model some priors and playing with the Bayes' theorem. The message is what experimentalists intuitively do: *when you have just a few observations, what you already know is more important than what the standard deviation of the data teaches you.*

Some will probably be worried by the arbitrariness of this conclusion, but it has to be remembered that: an expert can make very good guesses in his field; 20, 30, or even 50 % uncertainty in the uncertainty is not considered as significantly spoiling the quality of a measurement; there are usually many other sources of uncertainty, due to possible systematic effects of unknown size, which can easily be more critical. I am much more worried by the attitude of giving up prior knowledge to a mathematical convenience, since this can sometimes lead to paradoxical results.

4.4. Uniform prior for μ, λ and p in routine measurements

I find, on the other hand, that for routine applications the use of the uniform distribution for the center parameter of the normal distribution, usually associated to the true value, is very much justified. This is because, apart from pathological situations, or from particular cases in frontier research, even if one does not know if the associated uncertainty will be 0.1, 1, or 10 %, the prior knowledge on μ is so vague that it can be considered uniform for all practical purposes. The same holds when one is interested in λ of a Poisson distribution (counting experiments) or to p of the binomial distribution (measurements of proportions), under the condition that *normal approximation* is roughly satisfied, which is a kind of desideratum for the planning of a good routine experiment (otherwise it becomes a non-routine one). Taking into account the fact that, for routine measurements, the difference between mode and average of the final distribution is much smaller than $\sigma(\lambda)$ or $\sigma(p)$, we «recover» maximum likelihood results, but with a natural, i.e. subjective, interpretation of the results. This corresponds, in fact, to the case where the intuitive «dog-hunter probability inversion» (D'Agostini, 1998, 1999) is reasonable. For example, indicating by x the number of observed events in the case of Poisson distribution or of successes in the case of the binomial one, with the number of trials of the latter indicated by n, we get, simply

$$\lambda \sim \mathcal{N}(x, \sqrt{x}) \tag{1}$$

$$p \sim \mathcal{N}\left(\frac{x}{n}, \sqrt{\frac{x}{n}\left(1 - \frac{x}{n}\right)\frac{1}{n}}\right), \tag{2}$$

where $\mathcal{N}(\cdot, \cdot)$ is short hand for normal distribution of given average and standard deviation.

The recommendation I usually like to give, to check «a posteriori» if the uniform prior is suitable or not is the following: first evaluate central value and standard deviation according to the approximations (1) and (2); then try to judge if the central value «disturbs» you, and/or the standard deviation seems to be of the order of your prior vagueness; if this is the case, it is now that you need to model down some priors, which will actually affect the posteriors; otherwise, priors will have no appreciable effect and the approximated result is good enough.

This «a posteriori» consideration of priors might seem questionable, but I find it absolutely consistent with the spirit of subjective probability. In fact, the priors one has plug into Bayes' theorem should reflect the status of knowledge as it is felt to be by the subject who performs the inference. But sometimes it can be difficult to model this information consciously, or it might simply take too much time. The comparison of the approximate result got from a uniform prior with the result that the researcher was ready to accept can help, indeed, to raise this status of prior knowledge from the unconscious to the conscious.

5. PRIORS FOR FRONTIER SCIENCE

The question is completely reversed when one is interested in quantities whose value might be at the edge or beyond the sensitivity of the experiment (perhaps even orders of magnitudes beyond it) and if the quantity itself makes sense at all. This is a typical situation in particle physics or in astrophysics, and it is only to these kind of measurements that I will refer to as «frontier science measurements». However, even though they are «frontier», most of the measurements performed in the above mentioned fields belong, in fact, to the class of «routine measurements».

I would like to illustrate this new situation with a numerical example. Let us imagine that an experiment has been run for one year looking for rare events, like magnetic monopoles, proton decays, or gravitational waves. The physics quantity of interest (i.e. a decay rate, or a flux) is related to the intensity r of a Poisson process. Usually there is also an additional Poisson process to be considered, associated with the physical or instrumental background which produces observables indistinguishable from the process of interest (r_B). The easy, although ideal, case is when the background is exactly zero and at least one event is observed. This case prompts researchers to make a discovery claim. Let us consider, instead, the situation when no candidate events are observed, still with zero background. The likelihood, considering 1 year as unit time, is $f(x = 0\,|\,r) = e^{-r}$. Considering a uniform prior for r, we get $f(r\,|\,x = 0) = e^{-r}$ (see figure 1), from which a 95 % probability *upper limit* (r_u) can be evaluated. This comes out to be $r_u = 3$ events/year and it is a kind of standard way in HEP of reporting a negative search result[8]. The usual interpretation of this result is that, if the process looked for exists at all, then there is 95 % probability that r is not greater then r_u. But, I find that often one does not pay enough attention to all the logical implications contained in this statement, or in all the infinite probabilistic statements which can be derived from $f(r\,|\,x = 0)$. This can be highlighted considering statements complementary to the standard ones, especially in those cases in which the experimenters feel that the detector sensitivity is not suitable for searching for such a rare process. The embarrassing reply to questions like *«do you really believe 5 % that* r *is greater than* $\mathrm{r_u}$*?»*, or *«would you really place a 1 to 19 bet on* $\mathrm{r} > \mathrm{r_u}$*»* shows that, often, $f(r\,|\,x = 0)$ does not describe coherent beliefs. And this is due to the fact that the priors were not appropriate to the problem. For example, a reasearcher could run a cheap monopole experiment for one day, using a 1 m^2 detector, find no candidates and present, without hesitation, his 95 % upper limit as $r_u = 3$ monopoles/m^2/day, or 1095 monopoles/m^2/year. But he would react immediately if we made him aware that he is also saying that there is 5 % chance that the monopole flux is above 1095 monopoles/m^2/year, because he *knows* that $\mathcal{O}(1000)$ m^2 detectors have been run for many years without observing a convincing signal.

[8] It is worth noting that many physicists are convinced that the reason for this value is due to the fact that the probability of getting 0 from a Poisson of $\lambda = 3$ is 5 %. This is the classical arbitrary probability inversion (3) which in this case comes out to be correct, assuming a flat prior, due to the property of the exponential under integration.

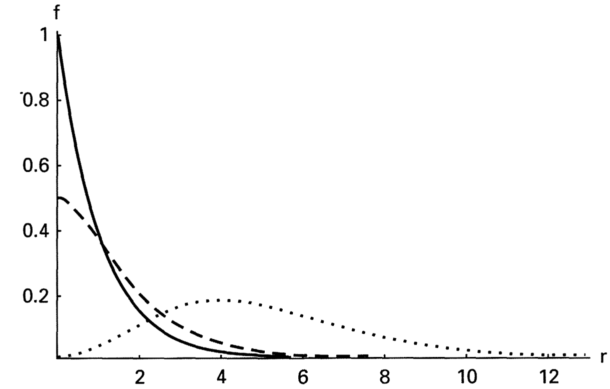

Figure 1. Final distribution for the Poisson intensity parameter r, obtained from a uniform prior and with the following values of expected background and observed events: 0, 0 (continous); 1, 1 (dashed); 1, 5 (dotted).

The situation becomes even more complicated when one has a non zero expected background and a number of observed candidates superior to it. For example, researchers could expect a background of 1 event per day and obsere 5 events. Differently from the above example of the monopole seach, let usimagine that the prior knowledge is not so strong that all the 5 events can be attributed with near certainty to background. Instead, let us imagine that the experimenters are here in serious trouble: the p-value is below 0.5 %; they do not believe strongly that the excess is due to the searched for effect; but neither do they *feel* that the probability is so low that they can decide not to publish the result and miss the chance of a discovery. If they perform a standard Bayesian analysis using a flat prior they will get a final distribution peaked at 4 which looks like a convincing signal, since it seems to be well sparated from 0 (see figure 1). They could use, instead, a Jeffreys' prior and find no result, since $P(r \leqslant r_0)/P(r > r_0) = \infty$ for any $r_0 > 0$. It is easy to see that in such a situation pedantic use of the Bayesian theory («Prior, Likelihodd → Final») leads to an embarrassing outcome whatever one does.

Therefore, in the case of real frontier science observables, the best solution seems to be that one has to abstain from providing final distributions and publish only likelihoods, which are degrees of beliefs too, but they are much less critical than priors. But reporting the likelihoods as such can be inconvenient, because often they do not give an intuitive and direct idea of the power of different experiments. Recently, faced with problems of the kind described above, I have realized that a very convenient quantity to use is a function that gives the bayes factor of a generic value of interest with respect to the asymptotic value for which the experimental sensitivity is lost (if the asymptotic value exists and the Bayes factor is finite) (Zeuss, 1999; D'Agostini and Degrassi, 1999; Astone and D'Agostini 1999). In the simple case of the

Poisson process with background that we are considering, we have

$$\mathcal{R}(r) \equiv \frac{f(x)|r, r_B)}{f(x|r=0, r_B)} \tag{3}$$

The advantage of this function is that it has a simple intuitive interpretation of *shape distortion function* of the p.d.f. (or a *relative belief updating ratio*[9]) introduced by the new observations. As long as $\mathcal{R}$ is 1 it means that the experiment is not sensitive and the shape of the p.d.f. (and hence the relative beliefs) remain unchanged. Instead, when $\mathcal{R}$ goes to zero the beliefs go to zero too, no matter how strong they were before. Moreover, since the $\mathcal{R}$ differs from the likelihood only by a multiplicative factor, it can be used directly in the Bayes' formula when a scientist wants to turn them into probabilities, using subjective priors. Different experiments can easily be compared and the combination of independent data is performed multiplying the different $\mathcal{R}$'s.

The function $\mathcal{R}$ is particularly intuitive when plotted with the abscissa in log scale. For example, figure 2, shows the result in terms of the $\mathcal{R}$ function for the same cases shown in figure 1. Looking at the plot, one can immediately get an idea of what is going on. For example, it also becomes clear where the problems with the flat prior and with the Jeffreys' prior come from. We can also now understand which kind of priors the hesitant researchers of the above example had in mind. Their prior beliefs were concentrated some order of magnitudes below the peak of $\mathcal{R}$, but with tails which could also accommodate $r \sim \mathcal{O}(4)$. This is in agreement with the fact that after the observations the intuitive probability for $r > \mathcal{O}(1)$ becomes sizable (5, 10, 30 %?) and the researchers do not have the courage not to publish the result.

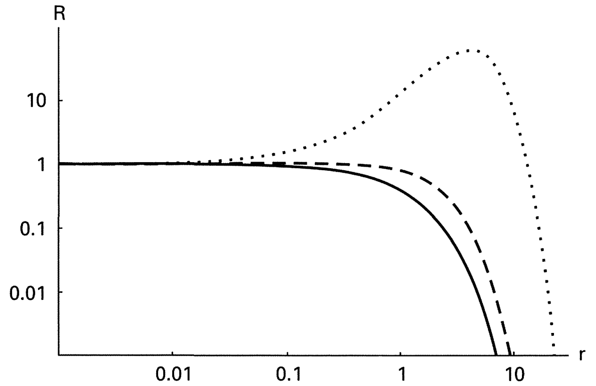

Figure 2. Bayes factor with respect to $r = 0$ for the Poisson intensity parameter r obtained from the following values of expected background and observed events: 0, 0 (continuous); 1, 1 (dashed); 1, 5 (dotted).

Finally, let us comment on upper (or lower) limits. It is clear now that, exactly in those frontier situations in which the limit would be pertinent, a highly intersubjective 95 % probability limit does not exist. Therefore one has to be very vareful in providing such a quantity. However, looking at the plots of Figure 2 it is also clear that one can talk somehow about a bound which roughly separates the region of «possible» values from that of «impossible» ones. One could then take a conventional bound, which could be the value of r for which $\mathcal{R} = 0.05$, or 0.5, or any other. The important thing is to avoid calling this conventional bound a 95 % or a 50 % probability limit. If instead one really wants to give a probability limits, one has to go through priors, which should be precisely stated. In this case, if I really had to recommend a prior, it would be the uniform distribution. This is not, only, for mathematical convenience, but also because it seems to me that it can do a good job in many cases of interest. In fact, one can see that it gives the same result as any other reasonable prior consistent with the positive attitude of the researchers which have planned and financed the experiment (for example, if an experimental team performs a dedicated proton decay experiment with the intention of making a good investment of public money, it means that the physicists really do hope to observe a reasonabe amount of signal events for the planned sensitivity of the experiment).

6. CONCLUSION

The key point expressed in this paper is that there is no need to «objectivise» Bayesian theory, treating subjectivism as if it were something we should be ashamed of. Only when this point is accepted and Bayes' theorem is correctly placed within the framework of subjective probability, with clear role and limitations, can the anxiety about priors and their choice be overcome. Once this is achieved, either we can choose the priors which best describe the prior knowledge for a specific problem; or we can «ignore» them in routine aplications, thus «recovering» maximum likelihood results, but with transparent subjective interpretation, and with awareness of the assumptions we are using; or we can decide that priors are so critical that only likelihoods or Bayes factors can be provided as the outcome of the experiment; or we can use the Bayes theorem in a reverse mode, to find, out which priors we had, unconsciously, that give rise to the beliefs we have after the new observation; finally there are some cases in which it is even more practical to skip the Bayes' theorem and to assess directly the degree of belief. With respect to this last point, I would like to remind the reader that, in fact, if one thinks that probabilities must only be calculated using the Bayes' rule, one gets trapped in an endless prior-final chain.

---

[9] In fact, in the case $f_0(r) \neq 0$ one can rewrite (3) in the following way

$$\mathcal{R}(r) = \frac{f(r|x, r_B)/f_0(r)}{f(r=0|x, r_B)/f_0(r=0)}.$$

As far as reference priors are concerned, they could, indeed, simplify the life of the practitioners in some well defined cases. However, their uncritical use should be discouraged. First, because they could lead to wrong, or even absurd, results in critical situations, if reference priors are preferred to case motivated priors just for formal convenience. Second, and more important, because of they might give the impression of dogmatism, which, together with the absurd results obtained through their misuse, could seriously damage the credibility of the Bayesian theory itself.

REFERENCES

1. Astone, P. and D'Agostini, G. (1999). *Inferring the intensity of Poisson processes at the limit of the detector sensitivity (with a case study on gravitational wave burst search), Tech. Rep.* CERN-EP/99-126.
2. Bayes, T. (1764). *An assay toward solving a problem in the doctrine of chances*, Philosophical Transactions of the Royal Society, **53**, 370-418. Reprinted in S. J. Press, *Bayesian Statistics: Principles, Models, and Applications.* John Wiley & Sons Ltd.
3. Berger, J. O. and Berry, D. A. (1988). Statistical analysis and the illusion of objectivity, *American Scientist* **76**, 159.
4. Bernardo, J. M. (1997). Non-informative priors do not exist, *J. Stat. Plan. and Inf.* **65**, 159-189. (With discussion.)
5. Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory.* John Wiley & Sons Ltd.
6. D'Agostini, G. (1999). *Bayesian reasoning in High Energy Physics - principles and applications.* Report CERN 99-03.
7. D'Agostini, G. (1996). *Bayesian reasoning versus conventional statistics in High Energy Physics,* Proc. of the XVIII International Workshop on Maximum Entropy and Bayesian Methods, Garching (Germany). (V. Dose, W. von der Linden, R. Fischer, and R. Preuss, eds.) Kluwer Academic Publishers.
8. D'Agostini, G. and Degrassi, G. (1999). Constraints on the Higgs boson mass from direct searches and precision measurements, *Eur Phys J. C.* 663-675.
9. De Finetti, B. (1974). *Theory of Probability,* J. Wiley & Sons.
10. Earman, J. (1992). *A critical examination of Bayesian confirmation theory, Bayes or bust? The MIT Press.*
11. Feldman, G. J. and Cousins, R. D. (1998). Unified approach to the classical statistical analysis of small signals, *Phys. Rev. D* **57**, 3873.
12. Galison, P. L. (1987). *How Experiments End.* The University of Chicago Press.
13. Hume, D. (1748). *Enquiry Concerning Human Understanding.* Abailable at /www.utm.edu/research/hume/wri/1enq/.
14. International Organization for Standardization (ISO) (1993). *Guide to the Expression of Uncertainty in Measurement,* Geneva, Switzerland.
15. Jeffreys, H. (1961). *Theory of Probability,* Oxford University Press.
16. Lad, F. (1996). *Operational Subjective Statistical Methods - A Mathematical, Philosophical, and Historical Introduction.* John Wiley & Sons Ltd.
17. Particle Data Group, Barnet, R. M., et al. (1996). Review of particle properties, *Phys. Rev. D* **54**, 1-794.
18. Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers.
19. Rodríguez, C. C. (1999) Confidence intervals from one observation, (Tech. Rep. available at/omega.albany.edu:8008/).
20. Savage, L. J., et al. (1962). *The Foundations of Statistical Inference: A Discusion.* Methuen.
21. ZEUS Collaboration and Breitweg J., et al. (1999). *Search for contact interactions in deep inelastic* $e^+p \rightarrow e^+X$ *scattering at HERA* (to appen), *Eur. Phys. J. C.*