ADVANCED STATISTICAL METHODS IN PHYSICS, Durham, March 2002

http://www.ippp.dur.ac.uk/statistics

The Post-Conference Version

Objectively derived default "prior" depends on stopping rule; Bayesian treatment of nuisance parameters is defended.

George Kahrimanis *

26 March 2002

Abstract

Several previous attempts (mine, too) for defining objective priors have been ineffectual, on account of both unsure derivation and implausible results. In a matter-of-fact approach, the existence of a default "prior" probability density is established in a special case: if the experimental error, as a random variable, is known, independent of the true value. This finding is extended to the generic case. The resulting "prior" is the same as the one proposed by V. Balasubramanian (1996), following Jeffreys' final approach. Consequently, prejudice can be eliminated in Bayesian analysis. This prior-free posterior also has a consistent frequentist interpretation. This assessment calls for a review of comparison of Bayesian to frequentist methods. Only the Bayesian method is suitable for obtaining results from atypical data sets. Still, the comparison between classical and Bayesian results can point out for us that the sensitivity of an experiment may need enhancement in a certain range. (Alternatively, one can compute a classical goodness of fit, in this way also testing the plausibility of the model.) Although useful, the classical approach has certain severe side effects, such as coupling of the background with the measured signal even if no events are recorded, and the counterintuitive lowering of upper limits in the presence of systematic uncertainties. Remedies have existed for years, though not yet endorsed by everybody: these side effects have been suppressed by means of a mixed approach, in which the background and/or the systematic uncertainties are treated in a Bayesian fashion. Such mixing is defended again here, with the rationale that normally our beliefs regarding systematic variables and the background are far from controversial (unlike beliefs concerning the estimated variables) therefore a Bayesian treatment is suitable.

http://users.hol.gr/~anakreon/defprior

*anakreon@hol.gr

1 Introduction

To keep this presentation short, I assume that the reader is familiar with the Bayesian procedure (*e.g.* D'Agostini, 1999); in particular, with the difficulty associated with the apparent freedom in deciding the prior pdf (prior probability density, in the continuous case; here called simply "prior"). In Prosper's (2000) succinct words, "if in fact your answer depends very much on the prior, then the conclusion should be that you have insufficient data to say anything sensible". However, as he, Lyons (2000), and others have noted, in the case of zero yield, the extraction of an upper limit for the event rate involves critically the choice of prior. The reader can check that the result is zero if the prior is a density uniform in the logarithm of the rate, or in the meantime, but it agrees with frequentist calculations if the prior is uniform in the rate. (This is also discussed by F. James, 2000 and Cousins, 1995.) Usually results are reported in terms of rate; this may explain why the corresponding parametrization is the preferred one, but I think that it is also perceived as the natural parametrization, albeit in a vague sense.

On account of the usual choice being merely a convention, an unforgiving critic may call for the abandonment of Bayesian analysis regarding upper limits, for the Neyman construction of confidence intervals, unless we settle the question of prior. If this sounds too radical, please consider on what grounds we expect to find a finite 90% Bayesian upper limit: just because we know that, for a range of hypothetical values of rate near zero, the null sample reserves the preponderance of probability ($>$90%). This is definitely "frequentist thinking".

So we have a compelling reason to review the problem of the prior. In this work, the scheme "prior $\times$ likelihood = posterior" is not applied if there is no reliable initial information, or if one decides to explore that possibility. We shall see that we can still obtain a posterior pdf, but it will not depend on a prior. This does not amount to abandoning the Bayesian approach; we only extend it to the case of null predisposition. The outcome is a probability density of the subjective kind, not a confidence interval.

Here is the basic idea. Consider a measurement of a single parameter or variable, such that for any trial we obtain just a numerical outcome, whose deviation from the true value (in the following, the term "error" will denote this difference, not the statistical uncertainty of the measurement) is known to follow a certain pdf, independent of the true value and of the data. (The choice of parametrization is important.) If there is no reliable preliminary knowledge to support a prior density, then the only thing we know is the pdf of the difference of the true value from the outcome. This is an objective pdf, which can be verified in a frequentist sense. Then a pdf for the true value, of the subjective kind, can be generated from the outcome and the pdf of the error. It is the result of the measurement in the absence of predisposition. The Appendix has a more detailed examination of the prior-less posterior pdf, both in the Bayesian and in the frequentist approach.[1] As we shall see, a prior-free posterior exists not only in the idealized conditions specified above, but also in a generic situation.

Note that Bayes' Theorem is not involved directly in the method introduced here, nevertheless the procedure happens to fit the description "prior $\times$ likelihood = posterior" (it will be demonstrated) though in appearance only. The density that takes the place of prior is called here

[1]For readers familiar with Fisher's fiducial probability: in the Appendix it is proven that in certain simple situations, even in the absence of a prior, a posterior probability can still be extracted from the experiment, and is the same as Fisher's fiducial probability in that case. Yet in other ways the present work is contrary to related assessments and conjectures advanced by Fisher.

"default prior". It is derived from the experimental conditions in a verifiable manner. Those conditions include the "stopping rule". A different experiment may correspond to a different default prior. To combine the results of two experiments, one must calculate a new default prior. The derivation of the default prior may be a difficult problem in many real situations, if the stopping condition of the experiment is debatable (say, if it has stopped for lack of money or because of a quarrel, or due to unforeseen circumstances). In such cases, usually an approximate derivation would be adequate.

For a quick inspection of the applied procedure regarding the default prior, see Sec. 3.6.

The existence of the default prior will be proven in this paper, but it is a separate matter to consider in what kind of problem it is appropriate to apply it. One such case would be a measurement for which there is some justifiable prior, but someone suspects (or considers the possibility) that the preliminary information might be fabricated – we shall examine such an example. To my mind, objectivity in data analysis amounts to a consistently skeptical attitude with regard to prior information.

The question what constitutes a justifiable prior needs to be addressed. My suggestion is that only uncontroversial priors should be permitted, like those obtained from plain estimates or from direct sampling measurements. Until now, in the absence of real preliminary information about the object of measurement, sometimes subjective inclination has been used to derive a prior probability density. This may be in principle admissible – in certain cases, and only in reference to a particular mindset or interest group. But often this inclination should be expressed in a different way, not in the way of a probability density. For instance, if it is based on predilection for a certain hypothesis, we should form a utility function, not a density, that is higher at the domain favoured by the hypothesis than elsewhere. If the inclination is based on unspecified observations, for example "life would be impossible if the value of this constant were less than …", then it is better expressed as coming from previous experiments, to be combined with the experiment we analyse, rather than as a probability density. We shall return to this matter.

The present approach differs from the related previous attempts, such as the Jeffreys-Jaynes priors, in not assuming the scheme "prior $\times$ likelihood = posterior" but deriving it from elementary considerations. The results are identical with those of a suggestion by V. Balasubramanian (1996), to define the "natural" prior (same as the last of priors supported by Jeffreys, 1961) as the density that is derived from the local metric associated with distinguishability of hypotheses. A short time before the presentation of this work I also noticed that Cousins (1995) mentions certain works (his references 23 and 35) which use the "information metric" (also lead to the same "prior" in the case of a Poisson rate, at least). In view of those previous works, the new thing in this work is the derivation of the information-metric prior from fiducial probability (in a restricted sense, not following Fisher).

1.1 Outline of the argument

The term "translational symmetry" here will apply to a measurement whose outcome is a single number, used as estimator, if the error is a random variable independent both of the true value and of the data.

Theorem A. In any measurement characterised by translational symmetry for the error, an objective posterior can be easily defined, to be used by skeptical bettors who do not trust preliminary information.

Proof of this theorem is in the Appendix. That is, one proof in the Bsyesian manner and one proof in the frequentist manner.

If we divide the objective posterior by the likelihood function, we obtain a default prior, independent of the data, which is invariant under translation; that is, a density uniform in the related parametrization.

Theorem B. In a situation that is not exactly like the above but differs by a small perturbation, the objective posterior does not disappear. The new objective posterior should differ from the unperturbed one by an amount proportional to the size of the perturbation (to the first order).

If we divide the objective posterior by the likelihood function, we obtain a default prior; it remains to be proven that this density is independent of the data. (It is shown in the comment after Theorem E.)

Assumption C (it may well be a theorem). Any situation of interest can be considered as the final product of successive small perturbations, starting with a situation belonging to Theorem A.

In this way we address the question of existence in general.

Theorem D. Any measurement of a continuous variable or model parameter, if repeated a large number of times, leads to a situation that is covered by theorems A and B in a special parametrization, so that we calculate a default prior for the repeated measurement; it is asymptotically independent of the number of repetitions. It is also independent of the data.

It remains to be shown exactly independent of the number of repetitions, even for small numbers: see theorem E.

Theorem E. The default prior for a single measurement is the same as that of a repeated measurement.

This was not obvious, because the default prior is not derived from subjective predisposition. (The combination of two different measurements is another matter.) Now we know that the prior we found in Theorem D applies also to a single measurement. Now we know that a default prior is independent of the data in any case.

Theorem F. The above approach is checked equivalent to V. Balasubramanian's (1996) (or Jefferys, 1961) suggestion to define the "prior" as the density that is derived from the local metric associated with distinguishability of hypotheses. His argument in short: a likelihood ratio between hypotheses h_1 and h_2 applies not only to these two hypotheses but also to all other hypotheses in their neighbourhoods, such that are practically indistinguishable (to a degree) from h_1 or from h_2.

1.2 Comments about uniform priors

A uniform density corresponds to the simplest and least informative function one can think of. Yet it is a density, consequently it depends on the choice of parametrization; uniformity is only half an answer to the problem "find an unprejudiced prior". Now let us look into a technical aspect of this approach. Apparently a prior density must be normalised or normalisable because it represents continuously distributed probability. Yet this requirement can be relaxed (Williams, 1988) because the Bayesian procedure involves a normalisation anyway, which will be adequate if the integration of the likelihood does not diverge. Also, though unnormalisable a prior density is still applicable in calculating relative odds; that is, in terms of conditional probabilities. Therefore the concept of an everywhere uniform density is consistent: it does not presuppose imposition of cut-off limits.

2 Insight from plain circumstances

2.1 Example "Lost Beacon"

Somebody has lost an object; he suspects that it fell off his backpack while he was walking on a certain trail. That trail is very nearly a straight line. To obtain a clue about the location of the object, he decides to take advantage of the fact that it is an automatic light beacon, pre-set to emit a flash at midnight. To monitor the flash, he climbs a watchtower at the beginning of the trail, along with some friends, all of them Bayesians, bringing a device adapted for the purpose. Using that device, they record the altitude angle (from the horizon), then use it to calculate the distance of the beacon from the watchtower.

Practically the only source of uncertainty in this measurement seems to be the random error in the measurement of angle. That error is estimated to follow a narrow probability density, the same at any angle.

Next, the operator of the measuring device calculates a (subjective) probability density for the location of the beacon. Someone proposes a betting scheme based on that density: to divide the trail into bins of equal probability, so that the beacon would have the same chance of being found in any of those bins. By the way, he asks what was the assumed prior probability density in the calculations. Here is the following dialogue.

- I have assumed a density uniform in the angle parametrization, restricted within the range corresponding to the trail, as seen from the watchtower.

- Can we, please, review your reasoning? I think that I have a different proposal.

- My experience tells me that under these conditions the error follows a nearly Gaussian probability density, in the sense that in a long series of comparisons against a more precise method, the error is found to obey the above density. (Note that when finally we locate the beacon we will also assess the error of this angle measurement.) This probability is objective, independent of any prior beliefs. In so far as we exclude any other considerations, this density also functions as a subjective probability. The corresponding prior density is uniform in terms of angle, due to the likelihood being proportional to the resulting probability density (one may say that the instrument was built with that specification). Now let us take into account that the source of the flash cannot be outside the range of angles corresponding to the beginning and the end of the trail. That is, all current bets would be invalidated if the beacon is found to lie outside the trail. The related conditional probability is formed by truncating the above probability. This is equivalent to truncating the uniform prior density (in the angle parametrization).

- I wonder why you have not taken into account that, before we thought of using the watchtower, we had arranged to bet according to a density uniform in the distance parametrization. Should we disregard our previous belief? I know that a frequentist would think so, but I cannot see why a Bayesian would!

- I thought that your previous betting arrangement had not been based on any kind of prior knowledge but rather on total ignorance, merely assuming that there is a physical object somewhere on that trail. Yet the mere existence of an object on the trail does not imply that the appropriate parametrization is necessarily linear in the distance. Nobody had given me any substantial reason for that choice.

- I suppose that you were not around when we discussed this matter. Considering all conceivable ways that a person can lose an object on this trail, it seems that this density is the average. This is the true prior probability density, regardless of the instrument used to record the flash.

- I see now! The prior density that is uniform in distance is as good as proved, because of what we know about the way the beacon was lost. I agree that the calculation must be revised, to assume a prior density uniform in distance.

- (A third person) I prefer the angle parametrization for defining the uniform prior. Your reasoning is fine, but I happen to doubt the presumed prior information. Someone might be trying to deceive the rest of us.

- Well... yes, it is your right to reject the presumed prior knowledge, because it might have been introduced with the purpose to mislead the bettors. There is not only one rightful prior density; it depends on your assumptions. Yet that is not to say that anything goes: if you are committed to certain assumptions then you are bound to the corresponding prior density.

- (The same third person) I agree. Assuming that I reject any opinion that is not verifiable by me, I am obliged to use the angle parametrization to analyse this measurement.

- Actually the angle parametrization is not exactly the objective one, if we include a minor effect from the fog. I have not yet finished my calculation of this correction, so let us put the matter off for the moment.

- Talking about provisional betting odds, reminds me of a special case among us. Arthur is expecting his wife to drive him back home as soon as she finds some free time – but after he goes away he will not be eligible to win the bet. Therefore it makes sense for him to bet on the nearest bin (which will be searched first), in case the beacon is found there before his wife arrives. I do not think that his case corresponds to a new prior – only to the multiplication of probability by his utility function (as if it were a likelihood function). With regard to prior probability, he may adopt either one of the two approaches we have been discussing.

- This is not the only case of peculiar betting. Isolde believes that she saw the flash from her car, when he was coming here, so that now she has an independent, though approximate idea where the beacon lies. She has not yet decided what to use as a prior probability density. Her experience amounts to a second angle measurement, from a different point. If she accepts the reasoning that supports the distance-parametrization for the prior density her calculations will be uncomplicated, but if she doubts it she will have to look for a parametrization appropriate for the “objective” combination of the two measurements.

2.2 Connection with experimental physics

The above situation has a special property. If we ignore or disregard all prior information about where and how the beacon is said to have been lost, the error follows a known probability density, in terms of angle. In other words, the measuring procedure alone determines a default prior probability (also a default parametrization) if we exclude any unverifiable information. A skeptical bettor adopts this prior probability. We may decide to accept part of the offered information, such as that the beacon is on the trail. Then the original prior density is modified (here, it is truncated). If one adopts all available information and if it amounts to a probability density then that is his prior probability, without any influence from the measuring procedure.

Note that in the measurement of a physical constant we have no access to information about how the constant has come to be what it is – nothing like the story of the beacon accidentally been lost along the trail. Therefore, in estimating physical constants, we must find and adopt the default prior density, at least when it makes a considerable difference. There is no freedom in the choice of prior probability density for physical constants; at most we may be relaxed about finding the exact default prior density if an approximation seems adequate. Any subjective predisposition that we want to take into account must enter not in the way of a specially tailored prior density, but either as a utility function (like a prior likelihood) or as a set of other experiments, whether real or gedanken, whose results will be combined with the present experiment.

3 Proof and derivation

3.1 Existence of default prior in a generic situation

The question arises whether a prior-free posterior exists always or only in special situations, those marked by a continuous symmetry. With a minute variation of the above example, say if we take into account the effect of the fog, the symmetry of the problem disappears – but does this perturbation annul the existence of a prior-free posterior? Then a skeptical bettor would be left without any basis for betting, which does not seem plausible. Any minor variation, like the effect of the fog, must introduce only a small correction to a skeptic's betting odds, not a radical change.

We have considered the effect of a static perturbing factor, the fog. A perturbing factor may even be of a random magnitude, provided that its effect can be adjusted to be arbitrarily small.

If we divide the prior-free posterior of the perturbed case by the perturbed likelihood, we get a density that could be called the default prior, except that we have not yet shown that it is independent of the data. This will be shown at a later point in this proof; it will not be assumed until then.

At this point let us presume without detailed proof that a generic case of one-parameter estimation can be thought of as a mutation of a case of translational symmetry. That is, starting with an imaginary case in which the pdf of the error is independent of the true value and of the data, we apply successive small perturbations until we reach the case of interest. In this way a prior-free posterior is guaranteed in any case. Along with it, the existence of a default prior (not yet shown to be independent of the data).

3.2 The default prior density for repeated measurements – Part A

The default prior is not a real prior. It does not represent subjective belief. Therefore it makes sense to ask whether the default prior density for a measurement repeated twice (the results combined) is the same as the default prior density for the single measurement. At a later stage of this work shall see that it is so, but until then we will not assume this property.

For example, suppose that we measure the same angle twice, using a cyclical bevel. The data of the combined measurement consist of a couple of numbers. We can take advantage of the rotational symmetry and argue that the default prior of the combined measurement cannot be anything other than a density uniform in the angle. In a less symmetric situation it is not so obvious that the default prior does not change when we duplicate a measurement and combine the results. We shall return to this problem.

3.3 What is the default prior probability in a generic case – Part A

We have concluded the existence of a prior-free posterior pdf in any case. The corresponding default prior is defined from the assumption "prior $\times$ likelihood = posterior", though it still remains to be shown that the default prior is independent of the data.

In a generic case the measurement error, as a random variable, is not fixed; for example, it may be a function of the true value, or may depend on the data. Yet, if we consider imaginary repetitions of the same experiment, we can construct a situation that asymptotically possesses a translational symmetry, like the special examples discussed above. That symmetry will guarantee that the default prior will be independent of the data.

If the number of repetitions is large enough the shape of the likelihood approximates a narrow Gaussian. We shall look for a parametrization in which these Gaussians are of the same width.

Let us limit our search, allowing only parametrizations in which the (plain) derivative of the default prior is uniformly bounded in absolute value. As we shall see, this limitation does not exclude the answer to our problem. If the number of repetitions is large enough, so that the width of the likelihood is narrow enough, then the default prior is rather constant in comparison, and the probability density that expresses the result can be taken as proportional to the likelihood. The default prior (of the combined measurement) becomes asymptotically irrelevant – except that it exists.

In this way we have contrived a situation having a symmetry similar to that of the simple examples. The measurement error is the difference of the true value from the maximum-likelihood estimate. This error is a random variable that does not depend on the true value, because by construction the likelihood is a Gaussian of a known width.

Now let us derive the related parametrization, also show that it is asymptotically independent of the number of repetitions. We shall deal only with the one-parameter estimation problem, leaving the higher dimensions for future work.

Here H will represent the parameter of interest (a random variable, in the Bayesian approach). The data is partitioned into a finite number of bins, indexed by i. After N measurements, the yield in each bin is y_i (such that $\sum y_i = N$). For any value h of the parameter H, the probabilities $\Pr(y_i|h;N)$ are known binomial distributions, each of mean value $\hat{Y}_{iN}(h)$. (Note that $\hat{Y}_{iN}(h)$ are proportional to N, and that $\sum \hat{Y}_{iN}(h) = N$.)

For the bins with $\hat{Y}_{iN}(h) \geq 10$ (say) the Gaussian approximation will be supposed adequate. The rest of the bins will be ignored in the following calculations. The resulting error approaches zero as $N \to \infty$.

Corresponding to any data set $\{y_i\}$, the log-likelihood of h is $\lambda \equiv \sum \ln(\Pr(y_i|h;N))$. In the Gaussian approximation, the log-likelihood can be written as $-1/2 \ \sum \ (y_i - \hat{Y}_{iN}(h))^2/\hat{Y}_{iN}(h)$. Accordingly, $\partial^2\lambda/\partial h^2 \equiv -\sum_i \ \hat{Y}'_{iN}(h)^2/\hat{Y}_{iN}(h) \ + 2\hat{Y}'_{iN}(h)^2(y_i - \hat{Y}_{iN}(h))/\hat{Y}_{iN}(h)^2 + \hat{Y}'_{iN}(h)^2(y_i - \hat{Y}_{iN}(h))^2/\hat{Y}_{iN}(h)^3 \ -\hat{Y}''_{iN}(h)(y_i - \hat{Y}_{iN}(h))/\hat{Y}_{iN}(h) \ -\hat{Y}''_{iN}(h)(y_i - \hat{Y}_{iN}(h))^2/2\hat{Y}_{iN}(h)^2$.

This second derivative is of interest at the point of maximum likelihood, h_m, considering that asymptotically the likelihood approaches the Gaussian shape. The mean and the variance of Y_{iN} are proportional to N (for fixed binning and parametrization), so that compared with the first term the other terms are negligible in the limit $N \to \infty$. In words, the effect of random fluctuations in the data diminishes, and the choice of parametrization and binning becomes irrelevant. Therefore in the limit of large N, at the point of maximum likelihood, the second derivative of the log-likelihood is $-\sum \hat{Y}'_{iN}(h_m)^2/\hat{Y}_{iN}(h_m)$, with diminishing fractional uncertainty (which arises from random fluctuations in the data and from the freedom in the choice of parametrization).

We are interested in finding a parametrization of H, such that the second derivative at the point of maximum likelihood can be considered as independent of the true value, subject only to random fluctuations and other asymptotically vanishing effects. This problem is trivial. In the rest of this subsection let us assume that it has been solved. For completeness, in Subsection 3.6 we shall outline the practical steps of finding such a parametrization.

In this way the width of the likelihood function will be in effect independent of the true value of the parameter, for large enough N. (That width can be so short that the variation of the default prior becomes negligible.) At last we have a situation in which the experimental error is a random variable independent of the true value, like in the plain examples we have considered – except for the statistical perturbation. In conclusion, the default prior corresponding to N repetitions of the measurement is a uniform density in the special parametrization (or any linear transformations).

The constancy of the width (asymptotically) of the likelihood is equivalent to the condition

$$(\forall h) \ \ (\partial/\partial h) \sum \hat{Y}'_{iN}(h)^2/\hat{Y}_{iN}(h) \ = \ 0. \tag{1}$$

This condition is independent of N, because the mean of Y_{iN} are proportional to N. That is, if N is large enough for the Gaussian approximation to be valid and the randomness to be negligible, the same condition will also be valid for any larger number of repetitions, for the specified parametrization.

This condition can be expressed without the explicit mention of data bins and repeated measurements, as an integral in the space of data. Let R be the random variable that corresponds to the result of a single measurement, and $p(r|h)$ the (known) probability density of r if H is h. The above condition is equivalent to

$$(\forall h) \ \ (\partial/\partial h) \int_{-\infty}^{\infty} dr \left[\left(\partial p(r|h)/\partial h\right)^2 / p(r|h) \right] \ = \ 0. \tag{2}$$

Already we have a method to derive a default prior density for a measurement that has been repeated many times. We have not yet shown that the same density is the default prior density for the single measurement. This will be shown in the following.

3.4 The default prior density for repeated measurements – Part B

If a measurement is characterised by a continuous symmetry, like when we measure length with a ruler or angle with a cyclical bevel, the default prior density cannot be anything other than a density of the same symmetry. This is so regardless of the number of repetitions of the same measurement. In a generic case, when there is no such a symmetry, we have to work harder to show that the default prior is independent of the number of repetitions.

Remember that we regard any generic case as a deformation of a symmetric case. Now we shall see that the property "the default prior is independent of the number of repetitions" is preserved to the first order under a perturbation. In this way the generic case inherits this property from the imaginary symmetric case.

Here is a proof. Suppose that we have a case for which the default prior is independent of the number of repetitions. We shall consider perturbations of this case. Let us call the former case C_0 and any perturbed case C_m, where m is the size (with a sign) of the perturbation, assumed small. In this way we have defined an one-parameter family of cases. As concluded above, if C_0 is repeated, the default prior of ${C_0}^2$ is the same as that of C_0, but we do not yet know whether doubling C_m also leaves the default prior unaffected. Let us call $D_m(h)$ the difference of the default prior of C_m^2 from that of C_m (the definition of "difference" needs some attention in case the default priors are unnormalisable). Assuming that an analytical expansion is valid, we obtain to the first order $D_m(h) = m\, g(h)$, for some function r. For instance, with the opposite perturbation we get the opposite effect.

Now let us consider doubling not only measurements of the one-parameter family defined by the perturbation, but also composite measurements, like $C_m C_n$. That is, we are interested in the difference of the default prior of $(C_m C_n)^2$ from that of $C_m C_n$; let us call it $D_{mn}(h)$. Again we assume that an analytical expansion is valid, so that to the first order $D_{mn}(h) = (m+n)\, g(h)$. For instance, if the composite measurement consists of two opposite perturbations, then the doubled measurement has the same default prior, to the first order.

The last expansion can be generalised, for composite measurements consisting of more than two measurements, like $C_m \ldots C_n$: $D_{m \ldots n}(h) = (m + \ldots + n)\, g(h)$.

Now we shall compare two cases. The one is C_m, for some small m. The other is $(C_{m/N})^N$, where N is a very large number. Both cases have the same first-order expansion of the effect we are interested in; therefore, the difference between the default prior of C_m^2 and that of C_m is the same, to the first order, as the difference between the default prior between $(C_{m/N})^{2N}$ and that of $(C_{m/N})^N$.

But the last difference is zero, because it involves large numbers of repetitions of the same measurement. (We have shown that the default prior of repeated measurements is asymptotically independent of the number of repetitions.) Consequently $D_m(h)$, the difference between the default prior of C_m^2 and that of C_m, is zero to the first order in m, therefore $g(h) \equiv 0$. That is, the effect we have been discussing in this subsection does not exist. The default prior of a generic case does not change if we repeat the measurement twice; the same can be shown for additional repetitions of the same measurement.

3.5 What is the default prior probability in a generic case – Part B

We now see that the default prior of a measurement does not change under repetitions, therefore it is the same as the prior which we have discussed in relation to a large number of repetitions of the measurement. We have seen that asymptotically it is independent of the data.

At last we can speak of a default prior which is derived from a measurement procedure, such that does not change when the procedure is repeated and the results combined. As it were, knowing the measurement procedure creates a subjective prior equal to the default prior. This is not so really; it seems nonsensical to bet before a measurement using the default prior (for example before an angle measurement). Still this observation makes it easier to remember how to apply the notion of a default prior.

3.6 The practical prescription

A parametrization in which the default prior is uniform is not (cannot be) unique, but is defined up to a linear transformation. In the case of continuous data, we should check whether our assumed parametrization of the estimated parameter is such that Eq. 2 is satisfied. If not, we must look for a new parametrization, corresponding to a function $g(h)$, for which Eq. 2 is satisfied. Therefore $\int_{-\infty}^{\infty} dr\ \left[\partial p(r|h)/\partial h\, / g'(h)\right]^2 / p(r|h)$ is independent of h. The general solution is

$$g'(h) = C \left[\int_{-\infty}^{\infty} dr\ \left[\partial p(r|h)/\partial h\right]^2 / p(r|h) \right]^{0.5} \tag{3}$$

for any non zero value of C; we can set $C = 1$, because linear transformations do not matter in this problem.

The default prior is uniform in the parametrization defined by $g(h)$, therefore in the h parametrization the default prior density is (unnormalised)

$$DP(h)\ \propto; \left[\int_{-\infty}^{\infty} dr\ \left[\partial p(r|h)/\partial h\right]^2 / p(r|h) \right]^{0.5}. \tag{4}$$

In a case with discrete data, the required parametrization is

$$g'(h) = \left[\sum \hat{Y}'_{iN}(h)^2 / \hat{Y}_{iN}(h) \right]^{0.5}. \tag{5}$$

The right side of the above equation is also proportional to the default prior density in the h parametrization.

3.7 Distinguishability as the natural distance of one hypothesis from another

Here is a brief account of Vijay Balasubramanian's (1996) suggestion, in simplified terms. A likelihood ratio between hypotheses h_1 and h_2 applies not only to these two hypotheses but also to all other hypotheses in their neighbourhoods, such that are practically indistinguishable (to a degree) from h_1 or h_2. We shall see that in effect this intuitively plausible opinion is equivalent to the default prior that was derived above. The formula was proposed by Jeffreys, 1961, but it should not be confused with "objective" priors he had introduced earlier (*e.g.* 1932).

Assuming that h_0 is the true value of H, how can we define the "average" likelihood of any other value h? We know that likelihoods do not add, but log-likelihoods do (when we combine experiments without correlations). Therefore it makes sense to add log-likelihoods of repeated experiments, then divide by their number to define the average log-likelihood of h, assuming that h_0 is true. This would be $\int_{-\infty}^{\infty} dr\, p(r|h_0) \ln(p(r|h))$.

But a likelihood function may be multiplied by any constant without any effect. To standardise our result, let us refer to the likelihood ratio of h compared to h_0: the standardised average of the log-likelihood is $\int_{-\infty}^{\infty} dr\, p(r|h_0) \ln(p(r|h)/p(r|h_0))$. It has been called the relative entropy of h with respect to h_0.

In the Taylor expansion of the relative entropy as a function of the difference $h - h_0$, the first non-vanishing term is of the second order: $-1/2 \int_{-\infty}^{\infty} dr\, [(\Pr(r|h) - \Pr(r|h_0)]^2 / Pr(r|h_0)$, or $-1/2 \int_{-\infty}^{\infty} dr\, [(\partial/\partial h) \Pr(r|h_0)]^2 / Pr(r|h_0)\ dh^2$. That is, a local metric in the space of hypotheses is defined. The density induced by this local metric is called the "natural" prior. It would be uniform in a parametrization in which the local metric is also uniform. This consideration leads to Eq. 2, therefore Balasubramanian's (Jeffreys') natural prior is the same as the default prior.

4 Consequences and paradoxes

Let us look into the problem of combining two experiments in a single analysis. If the preferred parametrization is the same in both experiments, then the default prior is common to them and there is no departure from the usual procedure. The general case presents some difficulty. Indeed changes in the stopping rule can have an effect, as we shall see in an example.

The dependence of a Bayesian result on the stopping rule is surprising. Until now only a frequentist analysis would take the stopping rule into account. Note that, as we shall see, in the usual case of a fixed-time counting experiment, the parametrization derived here does not coincide with the one common in current practice but (unlike previously proposed guidelines for objective prior density) it does not deviate very far from it. And we should not forget that the stopping rule is indeed irrelevant if we have a dependable prior.

The default prior is not about betting odds; it is only a mathematical construction that can be utilised in the place of a prior if there is no basis for forming a prior. Yet we have also seen that the default prior can be handled as a real prior, such that depends on the measuring procedure. That is, if we change plans for the measurement then that hypothetical subjective belief must be modified. Only if one forgets that there is no true subjective belief to be modified there is a paradox here. Still, this dependence reminds me of a Bohr's complementarity. In either outlook, some objective element of our understanding of the situation depends on the experimental setup.

Last, a few more words about the way one can include his subjective inclination in a statistical analysis. Prejudice may be based on facts which are unclear or difficult to analyse quantitatively, or may be based on plausible assumptions concerning situations that have not been realised yet. In either case the problem of introducing prejudice in the analysis seems to be a special instance of combination of experiments. A simpler case is the utility-induced subjective bias. As an example, let us think of a physicist who does not know or think much of superstring theories, therefore his short-term professional prospects are better served if the posterior probability density does not carry any strong indication of superstrings. If that was the only consideration, he would just set the utility function to zero outside the region he feels comfortable with, but he also takes care to avoid appearing too unreasonable. At any rate, he has the right to introduce his utility function, but that is not the same as introducing a prior density.

4.1 Default prior depends on the stopping condition

The rate of a Poisson process can be measured with a counting experiment. We will examine two idealised cases, different only with regard to the stopping condition. In the one, the duration of the experiment is decided ahead of the execution of the experiment. In the other case, the experiment is supposed to run until the yield reaches a pre-set number.[2]

4.1.1 Assuming that the duration is pre-set

In a fixed-time experiment, the probability of yield n, is a function of the rate r: $\Pr(n|r;T) = e^{-rT}(rT)^n/n!$. We easily find that

$$\sum_n [(\partial/\partial r)\Pr(n|r;T)]^2/\Pr(n|r;T) = 1/(rT) \tag{6}$$

if we take into account that the variance is equal to rT. Obviously Eq. 1 is not satisfied, therefore the default prior is not uniform in the rate. In a parametrization q(r) that would satisfy Eq. 1, the above sum would be constant:

$$(dr/dq)^2[1/(rT)] = C. \tag{7}$$

Therefore $dq = C_1 dr/r^{0.5}$, which implies $q = C_2 r^{0.5} + C_3$.

To extract an upper limit from an experiment with null yield, assuming no prior, we should integrate the likelihood using the default parametrization. That is, a 90% upper limit $r_{0.90}$ is defined by

$$\int_0^{r_{0.90}} (dr\,/r^{0.5})e^{-rT} \;=\; 0.90\int_0^{\infty}(dr\,/r^{0.5})e^{-rT}. \tag{8}$$

This is different from the formula in common practice, which does not involve the factor $r^{-0.5}$. Yet the integrals are not divergent. In particular, one can check that

$$\int^R (dr\,/r^{0.5})e^{-rT} \;\equiv\; \pi^{0.5}\,erf(R^{0.5}). \tag{9}$$

The 63.2% limit (corresponding to $1-e^{-1}$) is $0.41/T$ rather than $1/T$ of the common practice. The often used 90% limit is $1.35/T$ rather than the classic $2.30/T$. The 99% limit is $3.32/T$ rather than $4.61/T$. The results of current common practice (that is, assuming a prior that is uniform in r) are consistently more conservative than the results with the default prior.

We can also compare the Bayesian average of the rate in either approach. With the commonly used prior, uniform in r, an experiment that has run for time T and has yielded N events produces a posterior with average $(N+1)/T$. With the default prior for a measurement of preset time, the average of the posterior is $(N+0.5)/T$; the bias is halved. (The derivation is not difficult.)

[2] These two cases do not correspond to real experiments exactly. A real fixed-time experiment would end prematurely if its counter reached the device limit before the end of the pre-set time. An experiment of pre-set yield should also be given a time limit, to make sure that it is performable.

4.1.2 Assuming that the yield is fixed

In the case of an experiment planned to run until the yield reaches a pre-set number N, the probability density of obtaining the Nth event at time T is

$$\Pr_N(T|r) = e^{-rT} r^N T^{N-1}/(N-1)!. \qquad (10)$$

We check whether Eq. 2 is satisfied. Taking into account that $\int_0^\infty dT\, e^{-rT} T^n \;\equiv\; n!/r^{n+1}$, we obtain

$$\int_0^\infty dT\,[(\partial/\partial r)\Pr_N(T|r)]^2/\Pr_N(T|r) \;=\; N!/r^2. \qquad (11)$$

In this case, too, Eq. 2 is not satisfied. Note that the default parametrization in this case is different from that of the case of a fixed-duration experiment.

$$(dr/dq)^2[N!/r^2] = C. \qquad (12)$$

Therefore $dq = C_1 dr/r$, which implies $q = C_2 \ln(r) + C_3$.

5 Comparing Bayesian with Frequentist results

This section of the paper is only an outline, because the first part already seems too long. Besides, the topic is examined by several speakers in this conference; my views may change because of their arguments.

When I first confronted this subject, my impression was that the two schools are irreconcilable. Later I noticed that subjective probability, the main feature of the Bayesian procedure, is not incompatible with the notion of objective probability. Any known objective probability is also subjective, in the sense of generating betting odds. (A "calibrated" subjective probability.) Therefore frequentist confidence intervals are also meaningful for a Bayesian, even if he usually prefers to calculate a subjective probability density rather than a confidence interval. In plain words, a Bayesian can do everything that a frequentist does, but not*vice versa*. For example, Quantum Mechanical probability is frequentist, and also Bayesian. However, some Bayesians object to the consideration of imaginary repetitions of the experiment, which are essential in the construction of confidence intervals as well as in the definition of classical goodness of fit.

5.1 General considerations on "goodness of fit"

5.1.1 Goodness of fit as the basis of confidence intervals

The basic idea in the construction of confidence intervals is that, for each hypothesis, we group together the most typical data for that hypothesis, and then we check which of these groups contain the actual data. The matching hypotheses form the confidence interval.

The degree of being typical, the "goodness of fit" (in a general sense) can be reckoned in one of several possible ways; for instance, with reference to likelihood ratios, or to χ^2. It is a statistic of a (tentative) data set in reference to a hypothesis. This is not to say that any statistic used for appraising goodness of fit is also appropriate for the construction of confidence intervals. Some are outright inapplicable, such as a Kolmogorof-Smirnof test, with which we grade specific aspects of "goodness" only. On the other hand, the likelihood ratios used for the Neyman construction in the "unified approach" of Feldman and Cousins (1998) lead to puzzling results if they are used as plain indicators of goodness of fit (*e. g.* when we measure a nonnegative parameter, assuming that the error is Gaussian). Related questions are discussed in another presentation in this conference, by Fred James (2002-a).

If the data is such that the 99% (say) confidence interval lies outside the limits of sensitivity of the experiment, we have an indication that a new experiment is probably needed, designed to be sensitive at that range. Of course the result may have been caused by a random fluctuation; therefore we can just repeat the first experiment before we design a new one.

If the confidence interval is outside the physically allowed range (in other words the goodness of fit is low for all values in the physical range) we are faced with the problem that our data is atypical for any value of the parameter in the assumed model. If the model itself is doubtful, we will have a reason to look for some alternate explanation of the data.

We should note that the choice of statistic can affect the resulting confidence interval so much that, instead of finding an empty set inside the allowed range, a confidence interval is obtained, but outside the limits of sensitivity. For example, in the measurement of a mass which shows a negative result for m^2, central intervals can go totally outside the allowed range while the confidence intervals of the unified approach have a small portion inside. We see that the previous two paragraphs are meaningful only if the statistic for constructing confidence intervals has been carefully considered.

5.1.2 Goodness-of-fit test complementing Bayesian analysis

Every experiment has limited sensitivity, which we can assess with "what-if" trials. A (Bayesian) posterior probability density alone does not tell us what this limit is, and whether there is any indication that we may need to extend them with a new experimental design. Moreover, the posterior alone does not indicate whether the obtained data seem typical or irregular in reference to the assumed model. The latter property is needed in appraising the plausibility of the model itself, if its truth is in question.

The default prior does not deliver compensation for uneven sensitivity. On the contrary, it is proportional to the measure of distinguishability. (This is relative to the parametrization.) Consequently, the most intuitive way to display the posterior would be by plotting the likelihood in a parametrization with uniform default prior, because in this way we would also show the limitations of this experiment.

The question of whether the sensitivity of the experiment seems adequate and the question whether the model seems plausible can both be addressed with considerations involving the classical goodness of fit. Of course, we should first establish an appropriate way to define goodness of fit; that is, which statistic we adopt.

We also must define a goodness of fit that applies to a model, rather than to a specific value of the model parameter. Here I submit two alternative suggestions. The obvious first thought is to define it as the maximum goodness of fit in the allowed range of the parameter. This definition is the fairest to each model. It means that we use goodness of fit for parameter fitting (like when we find the least-squares fit) which is not always appropriate. We should not forget to take into account the uncertainty of systematic variables, like the total flux.[3] That is, we will vary the systematic variables, too, to find the combination that maximises the goodness of fit, not only for the main experiment but also for all prerequisite measurements, combined.

The above maximisation may involve complex calculations. It is far easier to calculate the goodness of fit at the "point" of maximum likelihood. In this way we easily take into account the systematic variables: first we calculate the full likelihood of the experiment (of both the model parameters and the systematic errors); then we multiply it with the likelihood of each independent systematic error; last, we find the maximum of the combined likelihood. This goodness of fit may be less than the maximum of course, but is much easier to calculate. I think that we should consider it as a candidate, at least as long as the calculated goodness of fit is sufficient.

The second choice is the right one if goodness of fit is assessed with a method not suited for parameter estimation, like the Kolmogorof-Smirnof test and related "empirical distribution-function" tests (Zech, 1995).

5.2 Effects of systematic errors in the calculation of frequentist intervals

When we calculate the frequentist 90% upper limit for a background-free experiment with zero yield, consideration of systematic error leads to a "physically unacceptable" result (Cousins and Highland, 1992; Zech, 1989). A larger systematic error produces a smaller upper limit, if we apply a consistently classical approach. The obvious remedy is to average over a probability density for the systematic error, that is, a Bayesian treatment of the error. A similar situation can be observed regarding the effect of uncertainty in the background, if the background is known approximately. (Gan, Lee, and Kass, 1998.)

In the following I shall argue that a consistent frequentist treatment of both the unknown, presumed decay rate and the known, if approximately, background does not reflect our understanding of the situation. The analysis should take into account our shared belief regarding the background. This can be done only with a Bayesian treatment.

Considering the frequentist procedure in terms of goodness of fit, a consistent frequentist treatment tests equally the assumption (model) of a Poisson decay rate and the assumption of background; if the experiment yields much fewer events compared with the expected background, the goodness of fit of the composite model is so bad that the confidence intervals shrink or even disappear.

But of course the intention of such an experiment is not to test the model of background. Everybody involved in the experiment or studying the results, all have a shared belief concerning the background. This belief must be expressed in the analysis, using a Bayesian treatment. Only if someone wants to test or challenge the common belief concerning the background, then the consistent frequentist treatment seems appropriate.

5.3 Zech's modification of frequentist procedure in the presence of known background

Suppose that are interested in setting an upper limit for a Poisson rate r in the presence of known background rate b, if the experiment has yielded N events. For confidence level α, the

[3] This would be essential if the experiment is very sensitive to the exact value of a systematic variable.

frequentist analysis amounts to solving for r the equation

$$1 - \alpha = \sum_{k=0}^{N} \Pr(k|r+b). \tag{13}$$

That is, any higher signal rate would make the probability of "yield $> N$" larger than α.

The result has several counterintuitive properties (Zech, 2001). One of them is that, if the yield is null ($N = 0$), the calculated upper limits for the rate of the signal depend on the expected background. This coupling is unacceptable, because the fact that the signal yield is zero is obvious, regardless of the expected background.

To avoid these problems, Zech (1989) has introduced a modified equation, which takes into account our knowledge that the background is not larger than N:

$$1 - \alpha = \frac{\sum_{k=0}^{N} \Pr(k|r+b)}{\sum_{k=0}^{N} \Pr(k|b)}. \tag{14}$$

This equation coincides with the corresponding Bayesian equation, assuming a prior uniform in r (Helene's equation).

This matter has been examined in several papers; I refer you to Cousins (2000) and Zech (2001). Zech's formula has been proposed again, independently, by Roe and Woodroofe (1999).

Cousins observes that Zech's equation would conform with frequentist coverage requirements for an ensemble of experiments conditioned to produce no more than N background events. Such a condition can be guaranteed only for simulated experiments. At any rate, the number N is the total yield, that is, it cannot be known in advance of a real experiment. Obviously this is not standard coverage, it is modified coverage, taking into account that the total background has to be no more than N.

If Zech's modification is the right approach, exactly what was wrong with the standard frequentist approach? To answer this question, note that in the case of a null yield, the confidence interval depends on the expected background, so that if the expected background is high the confidence interval is very small or disappears. This amounts to a rejection of the background hypothesis. Yet we usually want the analysis to test not our basic assumptions about the experiment, but rather the hypothetical model only – unless there is reason to doubt the validity of our underlying assumptions.

In the rest of this section let us see in what sense the modified approach amounts to a half-Bayesian, half-frequentist treatment. Instead of assigning standard Poisson probability to each tentative yield k ($\Pr(k|r+b)$, assuming the known b and a hypothetical r) we take into account that the yield k may have been generated in one of $\min(N+1, k+1)$ ways: if no event is from the background, if one event is from the background, and so on. That is, the probability is taken as the sum of $\min(N+1, k+1)$ independent Bayesian channels,

$$\Pr(k|r, b; N) = \sum_{n=0}^{\min(N,k)} \frac{\Pr(n|b)}{\sum_{i=0}^{N} Pr(i|b)} Pr(k-n|r). \tag{15}$$

This is equivalent to Eq. 10 in Cousins (2000) (after Roe and Woodroofe), and leads to Zech's equation (Eq. 14 above).

This treatment is Bayesian regarding the background, but frequentist regarding the signal. For that matter, note that the full Bayesian treatment can also be expanded in the same way; that is, $\Pr(k|r+b)$ is equal to that calculated in Eq. 15, since $k \equiv N$ in the Bayesian treatment.

6 Conclusions

The proposal to use information metric to define a "prior" is not new, but now we know that it can be derived from a simple assumption. This assumption seems plausible to physicists: when we obtain a reading from an instrument, and we know the distribution of the experimental error, then we already have a subjective belief about the measured quantity, without reference to what it is. This does not imply that we follow Fisher in any way he developed his idea of fiducial probability.

This approach is not an alternative to Bayesian inference (as Fisher insisted) but merely its extension in cases of null prior knowledge, or intentional suspension of belief. Besides, the scheme "prior $\times$ likelihood = posterior" is maintained, even if in appearance only.

It is a separate question when this method is the appropriate one, especially if there seems to be some prior available. There is no place for a prior in objective estimation of parameters for a physical model, because objectivity requires suspension of belief regarding the object of search. This prior may be perfectly acceptable in another context.

The default prior is not directly related to the question whether the sensitivity of the experiment is adequate for an approximate estimation, or does the real answer hide away from the range of sensitivity. The full result of a measurement is not only a "posterior" probability density, but also information about the sensitivity. Both would be conveyed by a plot of likelihood using a parametrization in which the information metric is constant. (Less distinguishable values would be plotted closer together than more distinguishable values.)

The choice of an approach (classical or Bayesian) and the criterion for assessing goodness of fit require discrimination: "finding the method with the properties appropriate to the way the results will be used or interpreted", in the words of F. James (2002-b). In particular, the pure Neyman construction tests in the same way our underlying assumptions about the experiment as the hypothesis of the model whose parameters we want to estimate. Therefore, systematic effects should be treated in a Bayesian fashion (after Zech and Cousins-Highland), unless there are doubts about basic aspects of the experiment.

Because there is no inherent conflict between Bayesian and frequentist methods (from the Bayesian point of view) calculation of the classical goodness of fit is essential in assessing the plausibility of the model in reference to the obtained data, after a Bayesian parameter estimation. Related methods are discussed in another contribution to this conference, by Aslan and Zech (2002).

Acknowledgements

The presentation of this paper was partly funded by the Onassis ($\Omega\nu\acute{\alpha}\sigma\eta\varsigma$) 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 that have lead to this work.

APPENDIX. Logical steps of deriving the prior-free posterior in the easiest cases

The existence of a prior-free posterior is shown here in detail, if the situation matches this description. The measurement produces just one numerical outcome, which is used as the estimate; no other plausible estimate is available; the experimental error (that is, the difference from the true value, not the uncertainty of the measurement) is a random variable independent both of the true value and of the data. This assumption applies in certain plain situations, like when we measure length with a ruler. It is easy to check that, if we change the parametrization, this property is preserved only with linear transformations.

If we have no prior knowledge about the true value, or if we want to put aside any predisposition, then there is no basis for applying Bayes' theorem. Yet the measurement leads to a subjective pdf for the unknown true value, as we shall see. This proof is very simple technically, but I think that it requires first to clarify the context of the problem using an example, such as in Section 2.1 or the example offered here.

A technician has calibrated an electronic thermometer. The instrument is a black box for him. He has noticed that the outcome is subject to some small random error, and has also formed a table with the probability distribution of the deviation of the reading from the true value. This error seems uncorrelated to the true value, also to the outcome. Then the technician learns that this instrument has been used in a previous measurement and has produced a reading r. He is not told what that measurement was about, nor is he is inclined to guess. Nevertheless his experience with this instrument leads him to apply a subjective pdf to the deviation of r from the true value, even without a prior.

Odd yet true, this result can be understood in either a Bayesian way or a frequentist way, as we shall see next.

The experimental error is defined as the difference of the output R from the true value T: $E = R - T$.[4] One can run tests or simulations, varying the "true value", to make sure that the distribution of the error E is independent of the true value and of the data. Therefore the error is a random variable of objective, verifiable probability density $f(E) = \Pr((T + E)|T)$.

Note that this property defines a set of parametrizations of T (the same for Y), which are related by linear transformations. In the "Lost Beacon" Example (Section 2.1) such a parametrization is in terms of angle (rather than in terms of distance).

The frequentist way

Consider the interpretation of the results of an HIV test. Usually we think of this problem in Bayesian terms, taking as prior probability the relative incidence of the infection in a group of people in which the subject arguably belongs. In the frequentist interpretation there is no sence in a posterior probability for this subject: he is either infected or not. Yet the Bayesian result is

[4] In a strict frequentist interpretation of that previous measurement, the true value is not a random variable but a fixed value t. After the measurement, the reading is a fixed value r. In that sense, the error is also not a random variable any more. Yet we shall see that the frequentist approach is rather more subtle than that.

meaninful also in the frequentist approach, in the sense of relative HIV incidence in a particular ensemble of people (F. James, 2002-c).

Similarly, for any muon trigger in AMANDA (Hill and DeYoung, 2002) we find the likelihood of a muon going up (rather than down), and we multiply it with our estimate of up-to-down ratio (our prior for "up") to obtain the Bayesian result. In terms of appropriately defined ensembles, this is also the probability in the frequentist sense (Cousins, 2002) though not exactly in the sense of the probability of *this* event being an up-going muon. The difference between the two approaches is not in the result, but in the interpretation.

After the probability of HIV infection and the up-going muons, let us consider the probability of the experimental error being between any two fixed values e_1 and e_2. In the strict frequentist sense, the error e after *this* measurement is no longer a random variable, but in terms of appropriate ensembles the probability can be defined (as for the HIV test), for example using subsets of the calibration data.

Therefore the pdf of the experimental error, $f(E)$, makes sense even in the frequentist approach, albeit only in terms of relative frequency in hypothetical runs of the same experiment. The resulting pdf for the measured variable T is

$$g(t) \equiv f(r - t) \tag{16}$$

We have obtained a frequentist interpretation of the fiducial probability $g(t)$. Note that it makes sense only if the "pivot" parameter E is a random variable of known distribution $f(e)$ for a fixed t, the same for any t.

The Bayesian way

For the sake of the argument, let us consider what happens when one takes into account his prior for T, $\pi(t)$ (assuming that a prior exists). The posterior $p(t)$ for the true value is
$p(t) \propto \pi(t) \, \Pr(r|t) = \pi(t) f(r - t)$.
Although f has been defined as a density, here it is proportional to the likelihood function; the correspondence is established with reference to a parametrization that realises the translational symmetry we have assumed. Note that the corresponding posterior pdf for the error is

$$\Pr(e|r;\pi) \propto \pi(r - e) f(e) \tag{17}$$

But if there is no prior for T, or if it is not admissible in the analysis of this experiment, the subjective pdf of the error remains $f(e)$ even after the measurement, because there is no intermediate step in Bayesian inference. In other words, note that in Eq. 17 the prior can be taken to be the pdf $f(e)$, our prior expectation for the error, and the corresponding likelihood is $\pi(r - e)$ (the reader can justify this approach in detail).

Therefore the resulting pdf for T is the fiducial probability

$$g(t) \equiv f(r - t). \tag{18}$$

We now have a Bayesian interpretation of fiducial probability. Note that in the professed refutation of the notion of fiducial probability (*e.g.* O'Hagan, 1994) a prior for T is taken into account, but this is contrary to the line of the above argument.

References

B. Aslan and G. Zech. (2002) Comparison of different goodness-of-fit tests. Contribution to this conference.

V. Balasubramanian. (1996) Statistical Inference, Occams Razor and Statistical Mechanics on The Space of Probability Distributions, Princeton University Physics Preprint `PUPT-1587`. Also available electronically as preprint `cond-mat/9601030`.

R. Cousins. (1995) Why isn't every physicist a Bayesian. *Am. J. Phys. 63,* 398. CERN scan `9501056`.

(2000) Additional comments on methods for setting confidence limits. Contribution to the Workshop on Confidence Limits at Fermilab, available as `http://conferences.fnal.gov/cl2k/copies/bcousins2.ps`.

(2002) Comments in the discussion after Hill and DeYoung (2002).

R. Cousins and V. Highland. (1992) Incorporating systematic uncertainties into an upper limit. *Nucl. Instr. and Meth. A 320,* 331-5.

G. D'Agostini. (1999) Bayesian Reasoning in High Energy Physics, *CERN Yellow Report 99-03*.

G. Feldman and R. Cousins. (1998) Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D 57* (7), 3873-89.

R. A. Fisher. (1935) The Fiducial Argument in Statistical Inference. *Annals of Eugenics 6,* 391-8 (1935)

K. Gan, J. Lee, R. Kass. (1998) Incorporation of the statistical uncertainty in the background estimate into the upper limit on the signal. *Nucl. Instr. and Meth. A 412*, 475-82.

G. Hill and T. DeYoung. (2002) Application of Bayesian statistics to muon track reconstruction in AMANDA. Contribution to this conference.

F. James. (2000) Two contributions to the Workshop on Confidence Limits at Fermilab, available as `http://conferences.fnal.gov/cl2k/fredjames_lectures.ps` and `http://conferences.fnal.gov/cl2k/copies/fjames.pdf`.

(2002-a) The relation of goodness-of-fit to confidence intervals. Contribution to this conference.

(2002-b) Comment on a paper by Garzelli and Giunti. Contribution to this conference.

(2002-c) Overview of Bayesian and Frequentist Principles. Introductory Lecture for this conference.

E. T. Jaynes. (1968) Prior Probabilities. *IEEE Trans. Systems Science and Cybernetics SSC-4*. Repr. Rao Tummala and Henshaw, eds. *Concepts and Applications of Modern Decision Models*, Mich. St. U. (1976) and in Jaynes, *Papers in Probability, Statistics, and Statistical Physics*, Reidel Publ. Comp., Dordrecht-Holland (1989).

H. Jeffreys. (1932) On the Theory of Errors and Least Squares. *Proc. Roy. Soc. 138*, 48-55.
(1961) *Theory of Probability* 3rd edition.

L. Lyons. (2000) Comments during the Workshop on Confidence Limits at CERN, posted at `http://preprints.cern.ch/yellowrep/2000/2000-005`.

A. O'Hagan. (1994) *Kendall's Advanced Theory of Statistics, Vol. 2B.*

H. Prosper. (2000) Presentation and comments during the Workshop on Confidence Limits at CERN, posted at `http://preprints.cern.ch/yellowrep/2000/2000-005`.

B. P. Roe and M. B. Woodroofe. (1999) *Phys. Rev. D 60*, 053009.

D. Williams. (1988) Comment on "Small-signal analysis in high-energy physics: a Bayesian approach". *Phys. Rev. D 38* (11), 3582-3.

G. Zech. (1989) Upper limits in experiments with background or measurement errors. *Nucl. Instr. and Meth. A 277*, 608.
(1995) Comparing statistical data to Monte Carlo simulation – parameter fitting and unfolding. `DESY 95-119, ISSN 0418-9833`.
(2001) Frequentist and Bayesian Confidence Limits. Draft available as `hep-ex/0106023`.