

# A Unified Approach to Understanding Statistics

F. James  
CERN, Geneva, Switzerland

Developing the Frequentist and Bayesian approaches in parallel offers a remarkable opportunity to compare and better understand the two.

## 1. TEACHING FREQUENTIST AND BAYESIAN STATISTICS IN PARALLEL

At the risk of hopelessly confusing my students, I have recently been teaching statistics with a “unified” approach, giving the Frequentist and Bayesian points of view in parallel. The goal of course is to enhance our understanding of both approaches, and I think that at least the teacher has learned something from the exercise.

The principal advantage becomes clear when we come upon a serious limitation or weakness in one approach, at which point we immediately see how the other approach handles (and often solves) this problem. The solution may come at the expense of accepting some other limitation or weakness not present in the first approach. The student then knows the trade-offs and eventually can decide which methodology he or she prefers for a given problem.

My motivation is the realization that if I want to solve all the problems generally considered to be statistical in nature, and I want to do that in the way most people do it, then both approaches are needed. If I limit myself to one approach, I can only solve a subset of problems in the appropriate way. This is justified below.

## 2. WHY DO WE NEED BOTH FREQUENTISM AND BAYESIANISM ?

I personally became convinced of this necessity when I realized the immense importance of two most successful statistical devices, one of which is Frequentist and the other Bayesian:

1. **Pearson’s Chi-Square Test.** This test is over 100 years old and I estimate it is used  $> 10^6$  times/sec on computers around the world. We know that theoretically it is one of the weaker parts of Frequentist statistics, since there can be no optimal Goodness-of-fit (Gof) test unless the alternative hypothesis is specified, but still the fundamental principle of converting a distance measure to a p-value has been so successful for so long that it cannot be dismissed out of hand. Physicists use it not only to accept and reject hypotheses, but we also use the fact that

we know the error of the first kind to correct counting rates, and we compare the distribution of p-values with that (uniform) expected under the null hypothesis to obtain further information used to calibrate the apparatus and to estimate background (errors of the second kind). It is hard to imagine doing physics without the Chi-square test. Statistics without the Chi-square test is like California without the automobile: some may consider it an improvement, but it’s not going to happen, so we had better learn to live with it.

2. **Bayesian Decision-Making.** If there is any statistical method used more often than Pearson’s Chi-square test, it is Bayesian decision theory. It is used not only by research workers and scientists, it is used (implicitly) by everyone every day. That is how we all update our knowledge about the world around us and that is how we make the hundreds of big and small decisions we have to make to live our everyday lives. Whether we realize it or not, most of our thinking goes along Bayesian lines, so this must also occupy an important place in our statistical toolbox.

But Gof testing is not allowed in the Bayesian paradigm because it violates the Likelihood Principle. And Bayesian reasoning is not allowed in the Frequentist paradigm because it requires subjective input. Since I wish to have both in my statistical toolkit, I cannot adopt either of the fundamentalist exclusive approaches, but must somehow allow elements of both.

## 3. DUALITY IN PHYSICS

Duality of this kind is well known in Physics. For centuries physicists argued about whether light was particles or waves. The argument was about the “nature of light”. It was assumed that light had to be of one nature or the other. Now it is known that in order to get the right answer to all problems, we need to use:

- wave formalism when the light is propagating,  
and

- particle formalism when it is interacting with matter.

The lesson: If your principles restrict you to only one formalism, you won't get the right answers to all the problems.

Here the trade-off is pretty clear. Methods based on Frequentist probability will be limited to repeatable<sup>1</sup> experiments. On the other hand, Frequentist probability is in principle independent of the observer, whereas Bayesian probability is as much a property of the observer as it is of the system being observed, so methods based on Bayesian probability are necessarily subjective (see 10.2 below).

## 4. BASIC CONCEPTS

I organize statistics in the traditional way:

1. Basic Concepts
2. Point Estimation
3. Interval Estimation
4. Hypothesis Testing
5. Goodness-of-Fit Testing
6. Decision Theory

This structure is important for classical statistics. Bayesian methods are more unified through their common use of Bayes' Theorem, so this separation of topics is not so important for the Bayesian side, but it should still be valid. The direct confrontation between the two methodologies for each of these topics is both interesting and revealing.

### 4.1. Probability

Nowhere is the confrontation more interesting than in the definition of probability, which is of course at the root of all methods. I distinguish three different kinds of probability:

1. **Mathematical Probability** is an abstract concept which satisfies the axioms of Kolmogorov. There is no operational definition (no defined way to measure it).
2. **Frequentist Probability** is defined as the limiting ratio of frequencies, which restricts its application to repeatable phenomena (see footnote below). It satisfies Kolmogorov's axioms, so it is a probability in the mathematical sense.
3. **Bayesian Probability** is a little harder to define because of the many different (and often vague) definitions found in the Bayesian literature, but I call it "degree of belief" and I use the operational definition based on the *coherent bet* of de Finetti, which also satisfies Kolmogorov's axioms.

### 4.2. Random Variable

This concept is very important for Frequentist statistics where one must know what is random and what is fixed. Random variables can take on different values when the identical experiment is repeated. Fixed variables always have the same value, even if the value is unknown.

In Bayesian statistics, on the other hand, experiments are not repeated and the concept of random variable is not needed. Instead, the important concept is "known" or "unknown". Probabilities can be assigned to the values of a quantity if and only if those values are unknown.

The result of the above is that the way of treating data and hypotheses is just the opposite in the two methodologies. Since the data is known, but random, Frequentists assign probabilities to data, but Bayesians do not [see below,  $P(\text{data}|\text{hyp})$ ]. On the other hand, since the truth of a hypothesis is not random, even if it is unknown, Frequentists cannot assign probabilities to hypotheses, but Bayesians do.

Some Bayesian authors refer to unknown values as "random", which is misleading, since even a physicist who uses Bayesian methods and assigns probabilities to different ranges of possible values of an unknown parameter would not consider the true value as random.

### 4.3. $P(\text{data}|\text{hyp})$

This probability is meaningful and important in both approaches, but is used in different ways depending on the approach. First we note that in the case where the data are the measured values of continuous variables, this is not actually a probability, but rather a probability density function (pdf). Then in the usual case where the hypothesis is the value of a continuously variable physical parameter (such as a particle mass or lifetime), this pdf is a function of both the data and the variable hypothesis, but it is only a pdf with respect to the data.

---

<sup>1</sup>In fact, it is not strictly necessary to be able to repeat the identical experiment, but only to have an ensemble of experiments analyzed with the same procedure

In Frequentist statistics,  $P(\text{data}|\text{hyp})$  is sometimes used as a probability (for discrete data) or as a pdf (for continuous data). That is when it is considered *as a function of the data*, and all possible data are considered, including the data not observed.

If the data are considered fixed (at the measured values), then this function is no longer a probability or a pdf, it becomes a *likelihood function*. It is sometimes used this way in Frequentist statistics, and *always* in Bayesian statistics. Fisher, who assigned the name *likelihood* to this function, was quick to point out that it was not a probability, since it does not behave mathematically according to the laws of probability.

Fisher's perspicacity was commendable, but unfortunately his advice has been lost in some quarters. Perhaps because of the way Bayes' Theorem (see 4.6 below) is normally written with all functions denoted by a  $P$  including the likelihood function, there is confusion in some writings between probability and likelihood.

The book of O'Hagan [1] is particularly sloppy on this point, as it confuses probability, probability density, and likelihood throughout. This results among other things in a large part of the book being devoted to the analysis of properties of the posterior pdf which are in fact arbitrary because they are not invariant under transformation of variables in the parameter.

#### 4.4. $P(\text{hyp}|\text{data})$

This probability is meaningful only in Bayesian statistics, where it is called the *posterior probability*, calculated using Bayes' Theorem. When the hypothesis involves a continuous variable parameter (the usual case), this is in fact not a probability but a pdf.

#### 4.5. $P(\text{hyp})$

This probability is again meaningful only in Bayesian statistics, where it is called the *prior probability*, a problematic input to Bayes' Theorem. As in the previous paragraph, this is usually a pdf, not a probability.

#### 4.6. Bayes' Theorem

This important theorem is a part of the mathematics of probability, so it holds for any kind of probability. Thus the use of Bayes' Theorem does not necessarily make a method Bayesian. It is rather the kind of probability appearing in the use of Bayes' Theorem that determines which approach is being used.

In practice, Bayes' Theorem is not used extensively in Frequentist statistics. On the other hand, it is the cornerstone of Bayesian statistics, the theorem which expresses  $P(\text{hyp}|\text{data})$  as the product of  $P(\text{data}|\text{hyp})$

and  $P(\text{hyp})$ . Used in this way, it is of course meaningful only in the Bayesian framework.

## 5. POINT ESTIMATION

The Fisher criteria for point estimates are properties of the sampling distribution of the estimates: consistency, bias, variance. This leads to the selection of the maximum likelihood (ML) estimator because it has optimal properties under rather general assumptions. Surprisingly, although the Fisher criteria are not invariant under transformation of variables in the hypothesis (the bias of the square of an estimator is not equal to the square of the bias), the maximum likelihood estimate **is** invariant.

For physicists, invariance is a very important property and we feel uneasy about methods that would give different answers depending on whether we estimate mass or mass squared. In particular, I would like to know if invariant equivalents of the Fisher criteria can be found. For example, if the bias were defined in terms of the expected median of the estimates instead of the expected mean, it would be invariant. Of course the median is more computationally intensive and it is not a linear operator, which was certainly an overriding consideration in Fisher's time, but that should no longer be an obstacle. Since the ML method is invariant, everything is fine, but it would be nice to understand better how an invariant method can arise from criteria that are not invariant.

On the Bayesian side, the criterion for point estimation is usually taken as maximizing the posterior belief, which if you use a uniform prior leads to exactly the same maximum likelihood solution as in the Frequentist case. However, in the Bayesian case the invariance is a fortuitous consequence of two particular choices, both of which are hard to justify and neither of which is invariant:

- The uniform or “flat prior”, which is of course only uniform in some metric, so it is not invariant, and there is not always any obvious preferred metric.
- Taking the maximum of the posterior pdf as the point estimate, which is also metric dependent. In fact, in the “natural metric”, the one in which the posterior density is uniform, there is no maximum. If we follow the recommendation that the metric to be used for the posterior should be the one in which the prior is uniform, that would seem to remove the arbitrariness from the procedure, but I don't know how to justify this recommendation except that it yields the ML result.

## 6. INTERVAL ESTIMATION WITH EXACT COVERAGE

The most commonly used methods for interval estimation are in fact only approximate, in the sense that they do not necessarily give exact coverage. The approximation tends to be good in the limit of large data samples, which is also the limit in which Bayesian and Frequentist methods give the same numerical results. Such methods, important as they are in practice, are not of much interest in this course, so I skip directly to exact methods, valid for small samples.

### 6.1. Obtaining Exact Coverage

The concept of coverage is of course very different for the two approaches:

- Bayesian coverage is the observer's degree of belief that the true value of the parameter lies inside the interval quoted. It is always exact in the sense that there is no need to make any approximations, but of course belief is not easily measured with high accuracy. The degree of approximation hidden in the prior density is seldom considered.
- Frequentist coverage on the other hand, is a property of the ensemble of confidence intervals that would result from repeating the experiment many times with the same analysis. Thus when we quote a 90 % Frequentist confidence interval, that doesn't mean that the resulting interval has 90 % coverage, but rather that the method produces confidence intervals 90 % of which will contain the true value of the parameter. Sometimes it may even be seen that a particular confidence interval is one of the 10 % that is wrong, an embarrassing phenomenon for the experimentalist who wants to publish that. On the other hand, the coverage can in principle be calculated to any desired accuracy and does not depend on the observer's beliefs.

Bayesian intervals come straight out of Bayes' Theorem as soon as the prior belief is specified, and Frequentist intervals with exact coverage can always be calculated using the Neyman procedure, but a few important problems may arise:

- **discrete data.** When the data are discrete (e.g., Poisson, binomial) the Neyman construction cannot be made to give exact coverage for all possible values of the unknown parameter. Then it is necessary to overcover for some values (usually all but a set of measure zero, in fact). This can be fixed, but only at the expense of making the interval depend on an extraneous

measurement or a random number. Bayesian intervals do not have this problem.

- **multidimensional data or hypotheses.** There is not much experience with the Neyman construction in more than two dimensions of data (normally the data are reduced to a statistic). The Feldman-Cousins variant [3] of the Neyman construction has been applied successfully to two-dimensional hypotheses, but we don't know much about higher dimensions. For Bayesian interval estimation, it is known that high-dimensional prior densities pose a very serious problem.
- **nuisance parameters.** Bayesians need only a prior for the nuisance parameters, then they integrate. This can be expensive computationally but is conceptually clean. For Frequentist intervals, various approximate treatments are possible, but it does not seem to be known theoretically how to obtain exact coverage for a large number of nuisance parameters. In practice it is tempting to add a pinch of Bayes and integrate over some density for nuisance parameters.

### 6.2. Choosing between Intervals with Exact Coverage

Coverage is not a sufficient criterion to determine confidence intervals unambiguously, so an additional criterion is needed. This is true for both Frequentist and Bayesian intervals.

- **Central intervals.** The most obvious solution is to take central intervals, with the same probability under each tail of the pdf. Frequentist intervals would be central in the data, but Bayesian intervals would be central in the hypothesis (the parameter being estimated). This means that a Frequentist central interval is not necessarily central in the parameter, and in fact may turn into an upper limit, whereas a Bayesian central interval must be two-sided, even when the data clearly indicate an upper limit only. On the other hand, Frequentist central intervals can be non-physical, whereas Bayesian central intervals must always lie in the allowed region for the parameter.

- **Most powerful intervals.** As central intervals have problems in both approaches, it is natural to look for a better criterion which will lead to "best intervals" in some sense. The sense is of course different for the two approaches, but the idea is essentially the same.

For the Bayesian approach, the obvious criterion is to accept the interval containing the values with the highest probability density. This

is in fact what is usually recommended, but we should note that this is not the same thing as choosing values of highest probability. Given a posterior pdf for a parameter, there is no way to define which values have highest probability, We know only the probability density, which is metric-dependent.

For the Frequentist approach, the situation is better. Even though there is no Uniformly Most Powerful range for the two-sided case, one can still apply the criterion suggested by optimal hypothesis testing, and choose the interval with the highest maximum likelihood ratio. It seems to be hard to find this construction described in the statistics literature, but it has appeared in the Physical Review [3].

- **Invariance of intervals under transformation of parameter.** It should be noted that the Neyman construction in general, and the Feldman-Cousins construction in particular are both invariant under transformations of variables in both the data and the parameters. Physicists are very fond of this property, since we know that any “true” theory has to obey some invariance principles including these. We are naturally uneasy about the Bayesian method of *highest posterior density* intervals. Of course it is always possible to argue that the posterior pdf contains all you could want to know about the parameter, and the interval concept is really a Frequentist invention not needed in Bayesian analysis, but still if Bayesian theory is “philosophically superior”, as O’Hagan claims [2], it ought to be possible to find an interval with good mathematical properties.
- **Non-physical (empty) intervals.** This topic provides a good opportunity to contrast the basic principles of Bayesian and Frequentist inference. Some Frequentist methods can yield measurements and even confidence intervals which lie entirely in the non-physical region. This is of course embarrassing since nobody really wants to publish such a value. One of the nice properties of Feldman-Cousins intervals [3] is that they cannot be empty (non-physical). On the other hand, this means they can be *biased* in the sense of hypothesis testing, whereas central intervals are not. So there is a trade-off between *bias* and *unphysicalness*, the old problem of measurements near a physical boundary. If a set of measurements near a physical boundary is unbiased, then some of them should lie on the wrong side of the boundary in order that the average be unbiased. In order to do that, some people have to publish values they know are wrong. For Frequentists, who think in terms of getting the

ensemble right, this poses no problem, but in the Bayesian approach, based on getting values you would want to bet on, it is stupid to propose a value that is wrong. In this case the duality is easy to understand: If you want to bet, use the Bayesian approach; if you want to get the ensemble right, use a Frequentist approach. Neither method is always right or always wrong. In practice people may seek a compromise like Feldman-Cousins, which has correct Frequentist coverage but avoids empty intervals at a minimum cost in bias.

## 7. HYPOTHESIS TESTING

Here we consider the testing of two simple hypotheses. If there is only one hypothesis, that is Goodness-of-fit testing, which comes in the following section.

The Bayesian approach produces the ratio of probabilities of the two hypotheses being correct. It can only give the ratio, because the individual factors in the ratio are not normalizable as true individual probabilities, but since the unknown normalization factor is the same, the ratio can be calculated. This ratio is exactly what is wanted in the Bayesian framework, since it is the *betting odds* which allows one to make an optimal bet for each data sample to belong to one or the other of the hypotheses.

The Frequentist method is based on determining the optimal way of classifying data in order to minimize the number of wrong classifications. Unfortunately, there are two ways of being wrong (the errors of the first and second kind), and the best one can do is, for a fixed probability of the error of one kind, to minimize the error of the other kind. Physicists refer to the error of the first kind as *loss* or *inefficiency*, and the error of the second kind is *contamination* or *background*. Traditionally one sets the level of loss that is acceptable, and the Neyman-Pearson lemma shows how to make the test with the lowest possible contamination.

It should be noted that there is no way to infer the error of the first or second kind knowing only the Bayesian betting odds, since the errors of the first and second kind require using the full sampling space and therefore violate the Likelihood Principle. Similarly, knowing only the errors of the first and second kind does not allow us to calculate betting odds, because that would require prior probabilities.

In this particular area, it seems to be especially difficult for people brought up on one approach to understand how the other one works. I know from my own experience, I had a lot of trouble to understand the meaning of the ratio of probabilities, since this combination does not normally appear in Frequentist methods. It does not allow you to calculate what physicists often want, which is acceptance and background, but

of course it is just what you want for betting. Similarly, the Bayesian literature contains statements (I am thinking in particular of Howson and Urbach[5]) that show confusion between acceptance and rejection of hypotheses on the one hand and decision theory on the other.

## 8. GOODNESS-OF-FIT TESTING

This is theoretically the weakest part of Frequentist statistics. There can be no optimal Gof test, because there is no alternative hypothesis, so it is impossible to define the power of the test. Gof is nevertheless the most successful part of Frequentist statistics, as pointed out earlier. The absence of an optimal test has been good for making work for those who like to invent new tests. I suppose it is this plethora of empirical methods which has given rise to the accusation by Bayesians that Frequentist statistics is *ad hockery*.

But if Gof is hard for Frequentists, it is even harder for Bayesians. The Likelihood Principle makes it impossible to do traditional Gof testing in the Bayesian framework, and attempts to do it in a proper Bayesian way are relatively recent and extraordinarily complex.

When I teach this, I simply say there is no Bayesian Gof test, which is in practice probably true, but of course the experts know this is simplifying things quite a bit. Since Bayesian Goffing requires an alternative hypothesis which should encompass all possible alternatives, the result is a very big hypothesis. This “hypothesis” is in fact a giant family of hypotheses containing additional unknown parameters, which requires a multidimensional prior in addition to the prior probability that this funny hypothesis is true.

I predict that James Berger’s virtuoso attempt to make a Bayesian Gof test (the Bayesian expression is *point null hypothesis*) will not be used by any physicist more than once. The proposed method has been published in two forms: The more accessible one in *American Scientist* [6] does not actually define the method, and the technical one is in *JASA* [7].

## 9. DECISION THEORY

This is the domain of Bayesian methodology. Bayesian decision rules are always best in the sense that for any non-Bayesian decision rule, there is always a Bayesian rule that performs at least as well. In addition, decisions are anyway subjective because loss functions are subjective. So it is really natural to use Bayesian methods for decision-making.

A simple example makes it clear why decisions should be Bayesian: Physicists are planning the next experiment or the next accelerator. The main goal is to find a new particle that is expected on theoretical grounds, but of course all the properties of this

particle (for example the mass) are not known. The detector can be optimized only for a certain range of masses, so a decision has to be made about the design of the detector. It is reasonable to optimize the detector for the properties we believe the new particle is going to have. That is a real prior belief. It may or may not be based on solid evidence or theoretical ideas. It is a mixture of knowledge and belief and intelligent guessing. It involves a real decision, namely there is money to build only one detector, and we have to decide how we are going to build it.

Once the experiment is performed, and let us assume that the new particle is found, the data also produce a measurement of the mass of the new particle. This step is no longer a decision, it is scientific inference. According to the scientific method, the measurement should be objective, not depending on the prior beliefs of the scientist doing the experiment. This is a good example to show the difference between a decision (which must depend on the subjective prior belief about the mass) and inference (which should **not** depend on prior beliefs).

## 10. SOME RELATED QUESTIONS

Now we are hopefully better prepared to address some more general questions.

### 10.1. Question: When is a calculation Bayesian?

Example: The Birthday Problem.

There are N people in a room. What is the probability that at least two of them have the same birthday. Assume all birthdays are equally probable and independent.

This is a standard problem in probability. Both Frequentists and Bayesians would get the same answer.

Variant of the Birthday Problem. In practice, there are small differences in the frequencies of occurrences of actual birthdays, so a variation consists in giving a (non-uniform) probability distribution for birthdays, or even correlations between them.

Again, Frequentists and Bayesians would make the same calculations, so there is still nothing particularly Bayesian here, although people with a Bayesian upbringing will view the input probabilities as *priors*, whereas for others they are simply the given conditions of the problem.

Final variant. You are put into a real room with 20 real people and asked to bet on

whether at least two have the same birthday.

NOW the problem becomes Bayesian, because now you must guess the (prior) probabilities and correlations between birthdays. It also becomes necessarily subjective, as real betting problems tend to be.

## 10.2. Question: Are Bayesian results necessarily subjective?

First of all, does it matter? In some cases, the answer is surely **yes**. Scientists want to be objective in reporting results. Bayesian fundamentalists may argue that complete objectivity is anyway impossible, so you better bring your subjectivity up front with a Bayesian analysis, still the typical physicist hesitates to introduce prior beliefs explicitly, and is especially unhappy to have to define probability as a degree of belief. In a recent course on Bayesian statistics at CERN, a young physicist in the audience commented: "But I can't publish my *beliefs*, no journal would accept that!"

As a result, many physicists' approach to Bayesian ideas is different from that of the statisticians. Many physicists think there is a "correct" prior, representing some kind of physical reality. Some even think of the prior as the distribution from which God randomly chose the values of the physical constants. Jaynes and Jeffreys tried to make Bayesian methods objective, arguing for example that two physicists with the same data and the same (lack of) prior knowledge should reach the same conclusions. I have the feeling that statisticians do not generally consider these efforts successful, but physicists tend to be less critical, probably because they would like it to be objective.

For me, the great advantage of Bayesian methods is to make the subjectivity explicit. He who tries to hide it (for example, with a flat prior), should ask whether he would not after all be happier with a method which doesn't need a prior at all.

## 11. WISH LIST FOR THE FUTURE

1. Scientists should learn both Frequentist and Bayesian statistics. The current situation is that most Bayesians seem to learn Frequentist statistics from other Bayesians (which is a disaster), and most people don't learn Bayesian statistics at all (which is equally bad).
2. I would like to see the Neyman interval construction and a Dinosaur plot (coverage as a function of the true value) in every new book or course on statistics. This topic is admittedly difficult for beginners, but even if the students

don't understand it, they will still learn (hopefully) that there is a unique Frequentist method for exact interval construction and they may get some grasp of what coverage means.

3. Give us a book on Bayesian Statistics which is mathematically rigorous (not pretentious, just correct) and does not make incorrect or unjustified statements about Frequentism. The best book I know is de Finetti [4], but he does indulge in some polemics, and the book has to be read with a pinch<sup>2</sup> of salt.
4. Investigate whether the Fisher criteria for point estimation can be made invariant under transformations of the parameter being estimated.
5. We should learn how to introduce nuisance parameters into the Neyman interval construction preserving coverage.
6. In some sense the ultimate problem is what to do when we wish to get Frequentist confidence limits on a counting rate, a typical classical problem except that one of the nuisance parameters is clearly Bayesian, for example its value is calculated by a theory in which we have a certain degree of belief. It appears that we must use both Frequentist and Bayesian probabilities in the same problem. Is there a way?

## 12. MY VIEW OF THE BAYESIAN-FREQUENTIST DUALITY

Example: Medical Research vs. Medical Practice

1. A research team investigates the effectiveness of different drugs in treating influenza. They are analyzing frequencies: how many people get better, how many do not, etc. They must be objective, and they must get the ensemble right. They use frequentist statistics and publish P-values, errors of the first and second kind, confidence intervals with exact coverage, etc.
2. A doctor has just finished reading the report of the above research team when a patient enters his office and complains of influenza-like symptoms. The doctor now uses Bayesian decision theory to decide how to treat the patient. He should find the best treatment for **this** patient, not for the ensemble of patients he might see. This is necessarily subjective, but based on the

---

<sup>2</sup>The book entitled *Theory of Probability* begins by stating that *probability does not exist*.

objective research of the research team. It is the doctor who wants to introduce any prior beliefs he has into his decisions; he does not want priors introduced already by the research team.

## References

- [1] Anthony O'Hagan, Kendall's Advanced Theory of Statistics, Vol. 2B, Bayesian Inference, Arnold (1994)
- [2] Anthony O'Hagan, *op.cit.*, p. 17
- [3] Feldman & Cousins, PRD 57 (1998) 3873-3889
- [4] B. de Finetti, Theory of Probability, J. Wiley (1974)
- [5] C. Howson and P. Urbach, Bayesian Reasoning in Science, Nature, 350, no. 6317, 371-374 (1991)
- [6] J. Berger and D. Berry, Statistical Analysis and the Illusion of Objectivity, American Scientist Vol. 76 p. 159 (1988)
- [7] J. Berger and T. Sellke, J. Am. Statist. Soc. Vol. 81, p. 112 (1987)