

# The proper fiducial argument

**Working Paper****Author(s):**

Hempel, Frank R.

**Publication date:**

2003

**Permanent link:**

<https://doi.org/10.3929/ethz-a-004526011>

**Rights / license:**

In Copyright - Non-Commercial Use Permitted

**Originally published in:**

Research Report / Seminar für Statistik, Eidgenössische Technische Hochschule (ETH) 114

# THE PROPER FIDUCIAL ARGUMENT

by

FRANK HAMPEL

Research Report No. 114

February 2003

Seminar für Statistik

Eidgenössische Technische Hochschule (ETH)

CH-8092 Zürich

Switzerland

# THE PROPER FIDUCIAL ARGUMENT

FRANK HAMPEL

Seminar für Statistik  
ETH Zentrum  
CH-8092 Zürich, Switzerland

February 2003

## Abstract

The paper describes the proper interpretation of the fiducial argument, as given by Fisher in (only) his first papers on the subject. It argues that far from being a quaint, little, isolated idea, this was the first attempt to build a bridge between aleatory probabilities (the only ones used by Neyman) and epistemic probabilities (the only ones used by Bayesians), by implicitly introducing, as a new type, frequentist epistemic probabilities. Some (partly rather unknown) reactions by other statisticians are discussed, and some rudiments of a new, unifying general theory of statistics are given which uses upper and lower probabilities and puts fiducial probability into a larger framework. Then Fisher's pertaining 1930 paper is being reread in the light of present understanding, followed by some short sections on the (legitimate) aposteriori interpretation of confidence intervals, and on fiducial probabilities as limits of lower probabilities.

## Keywords

Fiducial argument, fiducial probability, R.A. Fisher, foundations of statistics, statistical inference, aleatory probabilities, epistemic probabilities, structure of epistemic probabilities, upper and lower probabilities, frequentist statistics, axiom of frequentist epistemic probability, Bayesian statistics, intersubjective statistics, bets, odds, fair bets, successful bets, confidence interval, aposteriori interpretation of confidence intervals, Neyman-Pearson statistics, Behrens-Fisher problem.

## 1 Introduction

It may look surprising at first glance if at a conference on upper and lower probabilities a paper is presented about fiducial probabilities [7] which, whatever their detailed interpretation, belong to the class of proper probabilities. The reasons are as follows. (Proper) fiducial probabilities arose (somewhat surprisingly) as a special case in a side branch of my inference theory using upper and lower probabilities [19], and understanding of the general theory may help (and certainly has helped me) to understand the “mysterious” fiducial theory, and find an appropriate place for it in a larger framework. On the other hand, my experience both with the literature and with many oral discussions is that there still exists a lot of confusion, not only about the fiducial argument, but also about related concepts such as aleatory and epistemic probabilities, a frequentist interpretation of epistemic probabilities, and the difference between Fisher's and Neyman's interpretation of confidence intervals. Since all this is also part of my theory, an understanding of these

(historical) concepts is obviously required for a full understanding of my theory. But what is mysterious to me is that more than 70 years after Fisher's [7] first (and correct) paper about the fiducial argument, there is still no clarity about it, and most descriptions of it (following Fisher's erroneous later work) are from half true to plainly wrong and nonsensical. Therefore I thought it perhaps worthwhile to try to explain the (proper) fiducial argument and its surroundings in more detail.

The fiducial argument was meant to be a new mode of inference, making superfluous the appeal to a (usually unknown) apriori distribution to be entered into Bayes' theorem. From 1930 to about 1960, it was one of the "hottest" topics of debate in statistics, with participation of top statisticians across the whole spectrum, from J.W. Tukey to A.N. Kolmogorov. As some mathematical contradictions (within the later, false interpretation by Fisher) could be derived, the whole debate fell into oblivion soon after Fisher's death (in 1962), and many young statisticians today have never even heard of a fiducial argument or probability.

To give briefly one of the simplest examples: Let (entirely within a frequentist framework) a random variable  $X$  have the distribution  $N(\theta, 1)$  (normal with unknown location  $\theta$  and known variance 1), where  $\theta$  may be anywhere on the real line. Then for every fixed  $c \in \mathbb{R}^1$ , and for every (true, unknown, fixed) parameter  $\theta$  it is true that  $P(X \leq \theta + c) = \Phi(c)$  (with  $\Phi$  being the cumulative standard normal). Equivalently,  $P(\theta \geq X - c) = \Phi(c)$ . Now assume we have observed (a realisation of the random variable)  $X = x$  (e.g.,  $X = 3$ ). Then Fisher, using the "modus ponens", plugs in the observed  $x$  (this is the controversial "fiducial argument") and obtains  $P(\theta \geq x - c) = \Phi(c)$  (e.g., with  $c = 1$ ,  $P(\theta \geq 3 - 1) = P(\theta \geq 2) = \Phi(1)$  (this is interpreted as a "fiducial probability" for  $\theta$ , namely the probability that  $\theta \geq 2$ ; letting  $c$  move from  $-\infty$  to  $+\infty$ , one obtains the whole "fiducial probability distribution" of  $\theta$ ).

Now what is random?  $\theta$ ? But  $\theta$  was a fixed unknown constant. Moreover,  $P$  is actually  $P_\theta$  (this  $\theta$  is often suppressed). Are there two different  $\theta$ 's in the same formula?

In 1930 [7] and 1933 [8], Fisher gave correct (though cumbersome, brief and incomplete, hence apparently misunderstood) interpretations of this "tempting" result. But starting in 1935 [9], he really believed he had changed the status of  $\theta$  from that of a fixed unknown constant to that of a random variable on the parameter space with known distribution (cf. [11]). Apparently he needed this unfounded and false assumption in order to "solve" the Behrens-Fisher problem (the test for equality of means of two independent normal samples with unknown and possibly different variances, as opposed to the two-sample t-test). The "solution" was shown to be mathematically wrong; but Fisher was intuitively fully convinced of the importance of "fiducial inference", which he considered the jewel in the crown of the "ideas and nomenclature" for which he was responsible ([31], p. 370); and he vigorously defended his false interpretation up to his last statistics book [11]. Later on, most statisticians, unable to separate the good from the bad in Fisher's arguments, considered the whole fiducial argument Fisher's biggest blunder, or his one great failure (cf., e.g., [31], [6]), and the whole area fell into disrepute.

By contrast, I consider the (properly interpreted) fiducial argument the first (though highly incomplete) attempt to bridge the gap between a wholly aleatory Neyman-Pearson theory and a wholly epistemic Bayesian theory, either of which comprising only one-half of what statistics should be [20]; and Fisher does so by introducing (implicitly) frequentist(!) intersubjective epistemic probabilities (for a brief explanation of concepts, see Sec. 3). These concepts have strong implications for the everyday practical use of statistics, such as the aposteriori interpretation of confidence intervals (see Sec. 5). I thus agree with Fisher, not in his formal later interpretation of the fiducial argument (which is wrong),

but about the importance of the basic idea (and the correctness of his first interpretation, which he later denied).

I can only speculate about the reasons why the fiducial argument was not clarified earlier. Some reasons might be:

1. Lack of distinction between aleatory and epistemic probabilities (cf. Sec. 3). I believe Fisher felt the distinction intuitively, but he never clearly formulated it. For the Neyman-Pearson school, there exist only aleatory probabilities (very strictly so!), and for (strict) Bayesians there exist only epistemic probabilities (perhaps apart from a few simple cases where Hacking's principle might be applicable), hence the two school basically cannot even talk to each other (cf. also [20]).
2. Almost nobody seems to have checked on which probability space the (proper) fiducial probabilities can be defined! (Cf. the example above.) While the axiomatic foundation of probability spaces was done only a few years after Fisher's first fiducial paper [21], I find it surprising that apparently none of the later mathematical statisticians (with only one exception [25] known to me) has asked this basic question.
3. There seems to be an implicit conviction that there can be no frequentist epistemic probabilities (apart, perhaps, again from simple uses of Hacking's principle). This leaves only "logical" and subjectivist Bayesian results for scientists who really want to learn from data (and not just obey behavioristic rules), both of which are unsatisfactory in principle for them.
4. Within frequentist statistics, it seems often impossible for the thinking of statisticians to get away from the deeply entrenched paradigm of independent repetitions of the SAME experiment, although apart from a few applications, such as quality control and simulation studies, they hardly ever exist in science. Most scientists do DIFFERENT independent experiments each time, and frequentist properties of statistical methods can and should be evaluated with regard to such sequences of experiments. (If this should be considered an enlargement of the formal definition of "frequentist", then I find it long overdue. This point certainly was clear already to Fisher, for example, and other early writers.)
5. It is very tempting to believe that something which formally looks like a probability distribution is actually a probability distribution, without regard to the restrictions and interpretations under which it was derived. I am talking, of course, about the general misinterpretation of the "fiducial probability distribution".
6. Perhaps a major reason is Fisher's highly intuitive and condensed style of writing, which requires "reading from within" (trying to "feel" what he meant and thought) rather than "reading from the outside" (superficially taking words literally). - In addition, probably few statisticians went "back to the roots", to Fisher's first papers on the topic; it is the custom in our scientific enterprise to try to be always at the forefront of research; and the forefront in this case was leading astray because of Fisher's later blunder. (Still, the fact that some kind of blunder became known to exist, might have motivated a few more statisticians to study the origins more carefully.)

This paper contains some reflections on various reactions to the fiducial argument by other writers (Sec. 2), and mainly (throughout the paper, and specifically in Sec. 4) a

description of what I consider the proper fiducial argument, based on a new reading and interpretation of Fisher's first pertaining papers ([7], [8]) in relation to my more general approach [19]. It seems necessary to explain a few rudiments of my approach before Section 4 (Sec. 3). Some remarks on the dispute between Fisher and Neyman about the proper interpretation of confidence intervals are also added (Sec. 5), as well as a short section on fiducial probabilities as limiting and special cases of lower probabilities (Sec. 6).

The paper discusses only one-dimensional problems. The emphasis is on frequentist properties, as opposed to full conditionality and coherence; elsewhere I have shown that (symmetrical) optimal compromises between the two desiderata can be defined (including Bayes solutions closest to a frequentist interpretation, and vice versa), and that they can be numerically very close to each other ([17], [19]).

## 2 Some reactions to the fiducial argument

The history of Fisher's work on fiducial probabilities is excellently described by Zabell [31]. Briefly, Fisher discovered and solved the argument 1930 [7] for the correlation parameter, and 1933 [8] he solved it for the variance in normal samples (with a critical discussion of Jeffreys' approach to this problem). In 1935 [9], Fisher "solved" the Behrens-Fisher problem by assuming that the "fiducial distribution" is an ordinary probability distribution of a random parameter, and from then on he had to defend himself not only against lack of understanding, but also against justified criticisms. He tried - in vain - to escape these criticisms by some spurious conditioning arguments, but he was unable to retract gracefully from an untenable position (after previously having criticised Bayesians for the same reason). He even accepted the Bayesian derivation (with an improper prior) for the Behrens-Fisher test given by Jeffreys, after having criticised Jeffreys and stressing the differences of their approaches in 1933 [8]. His last major authoritative (though partly false) claims are in 1955 [10] and in his book 1956 [11]. (It should be clear that by the proper fiducial argument I mean Fisher's argument of 1930 and 1933, and not anything building upon his later erroneous work.)

It may be noted that already quite early, Fisher spoke occasionally and briefly of "fiducial inequalities" (in situations with discrete variables), thus faintly foreshadowing the use of upper and lower probabilities in these cases. In his 1956 book [11] and earlier, he contrasted fiducial probabilities and likelihoods as the main inference tools for continuous and discrete data, resp.; it seems to me that likelihoods might here better be replaced by upper and lower probabilities (while maintaining a central auxiliary role in both situations, of course).

By restricting himself to proper probabilities, Fisher obtained a rather limited theory. This reminds me of Bayes who in his scholium implicitly made the same restriction. Neither seems to have thought - or known - about upper and lower probabilities, although they had been introduced much earlier by James (Jacob) Bernoulli.

A number of statisticians, starting with Bartlett, tried to check the Behrens-Fisher test or to find conditions under which Fisher's new methods could be justified by other, objective arguments (cf., e.g., [1], [28], [29]). A number of other statisticians were more indirectly inspired by Fisher's work, trying to find something new in a similar direction. They include Fraser [12] trying to exploit group structures if they happen to be available, and especially Dempster ([4], [5]), whose work on different kinds of upper and lower probabilities led to the theory of belief functions by Shafer [27], Smets and others. Probably many if not most older participants at this conference have at one time or other thought hard about fiducial probabilities and were motivated by them for their own work, but

since these “offsprings” are not the topic of this paper, I apologize for not going into more details.

Kolmogorov [22] discusses fiducial probabilities in a summarizing report on contemporary British statistics. In Footnote 12, he suggests the introduction of a new axiom: if all conditional probabilities of an event, given the parameters, exist and are equal, then the unconditional probability exists and equals this value. At first, I was puzzled by this remark. Later, I thought that maybe this can be interpreted to be the axiomatic introduction of a new kind of probability (an epistemic one, to use present wording) which does not depend on any (unknown) parameters. Viewed this way, it may make deep sense, although it is still highly incomplete (for example, the underlying probability space and the epistemic interpretation are not discussed). - Incidentally, Kolmogorov quite naturally discusses sequences of different experiments (as opposed to repetitions of the same experiment); and he partly argues in favor of unconditional statements in applications, for practical reasons.

One of the greatest mysteries for me around the fiducial argument is why the extremely important 1957 paper by Pitman [25] seems to have gone virtually unnoticed in the discussion of the fiducial argument. (I owe the reference to Robert Staudte.) Pitman gives a lucid and deep mathematical description of the fiducial argument, what is right and what is wrong about it, and mathematical conditions under which it can be applied. He is not afraid of calling a mistake a mistake, such as Fisher’s [11] claim that the parameter has now the status of a random variable (simple acknowledgment of this fact would have made most past discussions of “fiducial probabilities” superfluous). In the same sentence (p. 325) he asserts that nevertheless “... we are able to make statements about the unknown parameter with a calculable probability of being right” (this is precisely my viewpoint). Pitman also discusses the fiducial distribution of several parameters, while making it clear that Fisher’s integrations for the “Behrens-Fisher solution” were simply not permissible. He does not go further into the conceptual interpretation, but any truly informed mathematical paper about the fiducial argument has to incorporate what he has to say. Nevertheless Pitman’s paper is not cited in Cox and Hinkley [3], Walley [30], Zabell [31], Efron [6], to name a few prominent works discussing fiducial inference; and this, although Pitman is not entirely unknown in mathematical statistics, the journal in which the paper appeared is not totally obscure, and the paper was an invited review article on Fisher’s most important book on foundations, and was written in Stanford, a place not totally provincial in statistics.

Cox and Hinkley [3] just give the usual “on the one hand ... on the other hand ...” type of discussion of the fiducial argument; and in exercise 7 of Ch. 7, p. 248 (accessible through the index), they try to discredit fiducial probabilities (without using the word here) by using an operation (curtailing the distribution) which was (correctly) explicitly forbidden in Fisher’s early works (though not in his later, incorrect works).

Most statisticians, after making up their mind that the whole fiducial argument was just a big blunder, tried to forget it. C.R. Rao [26] may have been one of the last famous statisticians to include it in an ordinary applied statistics textbook, though with reservations. But although the argument in its later form had been proven wrong, some statisticians, such as J.W. Tukey, still thought “there may be something to it”. I learned this attitude from him in the seventies, and in our last telephone conversation in July 2000, a few days before his death, he confirmed to me that this still was his opinion.

In his 1996 Fisher Lecture, Efron [6] gave (among other things) a reasonable sounding discussion of the desirability of something like fiducial inference in future statistics. But then (end of Section 8), out of the blue he suddenly speculates: “Maybe Fisher’s biggest

blunder will become a big hit in the 21st century!” I agree of course about the hit (though, slowly as statistics progresses, it may well be the 22nd century), but just by rereading his paper I can’t find any reason or justification for his optimism (unless there is a - direct or indirect - connection with my talk on foundations at Stanford in the Berkeley-Stanford Colloquium in March 1994). In any case, it is gratifying to see (also in the discussion of the paper) that some statisticians might be willing to take a fresh look at the fiducial argument, recognizing its basic importance for statistics.

### 3 Some rudiments of a new general theory of statistics

Before rereading Fisher’s 1930 paper, it seems necessary to explain some of the concepts which are hidden in and behind Fisher’s early work, and which should make this work more comprehensible.

The broadest view of my emerging theory is given in [14], some more solutions in [15], the probably most readable introduction in [18], and the most recent highly condensed survey in [19]. I first noticed the connection with fiducial probabilities in an abstract [13], and later in a chapter in [14].

Let me now try to explain my view of statistics.

I became more and more convinced that we have to distinguish between aleatory and epistemic probabilities. Aleatory probabilities (as in dice throwing) are supposed to be probabilities occurring objectively in Nature (in “random experiments”). (Of course, I am aware that the whole concept of an objectively existing Nature - as well as that of probabilities - can be criticized philosophically, but without it we could have no science, and we have been quite successful with it. Here I am trying to keep the discussion on a reasonably pragmatic level.) Usually, aleatory probabilities are unknown to us (except in simulation studies, or under Laplacean assumptions of symmetry), but we are trying to learn something about them. Aleatory probabilities are frequentist, that is, they obey the law of large numbers. This (besides the usual axioms) gives them an (approximate, but arbitrarily accurate) operational interpretation: In a long sequence of (usually different!) independent experiments, all with probability of “success” (different “successes”!) equal  $p$ , the observed fraction of “successes” will be close to  $p$  (with the usual precise mathematical formulation).

But many statisticians, and users of statistics, also want to learn something, and want to know some probabilities (and not only approximately, if it can be done). Probabilities which refer to our (personal or intersubjective, assumed or true!) state of knowledge are called epistemic. It would be nice if we could derive known epistemic probabilities from the unknown aleatory probabilities.

But the Neyman-Pearson theory considers only aleatory probabilities. Neyman ([23], [24]) explicitly excluded inductive inference, hence all learning processes and the epistemic probabilities which they could lead to. This is very satisfactory for pure mathematicians, because it keeps the basis of the theory mathematically and conceptually simple, and it may be tolerable for decision theorists; but it is frustrating for many applied statisticians and users of statistics who actually want to learn something from their data (also in a broader context), and not just “behave inductively” without any thinking being allowed.

On the other hand, all Bayesian theories (and we have to distinguish at least between Bayes - if we want to call him a Bayesian -, Laplace, Jeffreys, and the Neo-Bayesians, who may be further split up) use, at least in principle, only epistemic probabilities (except perhaps for the conditional distributions of the observations, given the parameters, which may be considered aleatory, but which are quickly transformed away). They start with



epistemic prior distributions for the parameters, and they end with epistemic posterior distributions or some predictions or decisions derived from them. Bayesian probabilities may be subjective (as with the Neo-Bayesians) or “logical”, “canonical” or “objective” (a very dubious term), as with the other Bayesian schools mentioned; these logical probabilities are intersubjective, that means, they are the same for scientists with the same data (and model) and the same background knowledge. But none of them has a frequentist interpretation (unless the prior chosen happens to be a true aleatory prior). The concept of a true, unknown aleatory probability distribution which governs the success of Bayesian claims and bets, is alien to (strict) Neo-Bayesian theory, and the self-assuring success of Bayesian “fair bets” results from them being evaluated by their own subjective priors. We shall see that Fisher, in effect, tried to introduce frequentist intersubjective epistemic probabilities (without using these qualifying terms).

It seems natural to describe epistemic probabilities by bets or odds (or odd ratios), as has been commonly done already centuries ago. Bayesian (pairs of) fair bets (“If I am willing to bet  $p : q$  on  $A$ , I am also willing to bet  $q : p$  on  $A^c$ ”) are two-sided bets and correspond 1:1 to ordinary probabilities (“ $P(A) + P(A^c) = 1$ ”). But if we are not in a decision situation, but in an inference situation, we may also consider one-sided bets, expressing partial lack of knowledge about the true probability (up to total ignorance); in order to avoid “sure loss” with bets both on  $A$  and on  $A^c$ , we must have  $P(A) + P(A^c) \leq 1$  and are thus led to (some form of) upper and lower probabilities. (The Bayesian claims for equality to 1 are circular and just not true, except for enforced decisions.)

If the expected gain of my one-sided bet, evaluated under the true (unknown) aleatory (not epistemic!) probability of  $A$  is nonnegative, I call my bet “successful”. Obviously, I cannot bet successfully on  $A$  (except  $0 : 1$ ) without any knowledge about its probability; but the amazing fact is that in general I can find nontrivial successful bets if I have independent past observations from the same parametric model.

The bets cannot, of course, in general be conditionally successful given any fixed past observation, and at the same time informative, because usually there may be totally misleading past observations; but they can be successful when averaged also over the distribution of the past observations. Their success can be operationally checked and empirically validated by considering long sequences of independent successful bets (from different experiments!); with bounded and sufficiently variable gains the average gain will most likely be  $> -\epsilon$  for  $n$  large enough.

My theory is mainly for prediction, because I find prediction in general more important in practice than parameter estimation (cf. [16]), and it can be checked empirically. But the theory can also be done for random (!) parameter sets; and then, in some cases, it just gives the (proper) fiducial probabilities (as frequentist epistemic proper probabilities).

## 4 Rereading Fisher’s 1930 fiducial paper

(It may be useful for the reader to get a copy of this paper, e.g. from [2].)

In the beginning of the paper, Fisher [7] attacks Bayes (incorrectly, but mildly) and the then Bayesians (to discuss this part would mean another section - there appear to be some parallels between the history of Bayes’ argument and of the fiducial argument); and then he discusses likelihood (describing it, like the fiducial argument, more as an empirical discovery rather than an invention). Starting at the bottom of p. 532 (p. 433 in [2]), he discusses the fiducial argument with the example of the correlation coefficient, even giving a table for  $n = 4$  which for every  $\rho$  gives the upper 95% value (now “confidence limit”) for  $r$ . And this relationship “implies the perfectly objective fact that in 5% of samples  $r$  will

exceed the 95% value corresponding to the actual value of  $\rho$  in the population from which it is drawn.” (And he goes on to define  $\rho_r$ , the “fiducial 5% value of  $\rho$ ” corresponding to a given observed  $r$ .) Thus the actual value of  $\rho$  (an unknown constant) may be anything (or have any prior distribution, for that matter): among all unselected(!) pairs  $(\rho, r)$  (typically from different experiments!), in 5% of all cases  $\rho$  will be smaller than  $\rho_r$ . The random event “ $\rho < \rho_r$ ” (where the randomness is in  $\rho_r$ !) has an epistemic (known to us), frequentist probability of 5%. We can bet on its truth successfully 1 : 19, and we can even bet on its complement successfully 19 : 1; that means, we have not only a lower, but a proper probability for this event and a Bayesian (pair of) fair bet(s). The bets can be validated by taking any sequence of (different) independent experiments with arbitrary values of  $\rho$ , the (unselected!) observed values of  $r$  and the corresponding values of  $\rho_r$ ; in exactly 5% of all cases (in the long run),  $\rho$  will be found to be less than  $\rho_r$ .

“If a value  $r = .99$  were obtained from the sample, we should have a fiducial 5%  $\rho$  equal to about .765. The value of  $\rho$  can then only be less than .765 in the event that  $r$  has exceeded its 95% point, an event which is known to occur just once in 20 trials. In this sense  $\rho$  has a probability of just 1 in 20 of being less than .765.” Here (if “in this sense” is interpreted correctly) Fisher is still correct, but dangerously close to his later mistake of considering the “fiducial distribution” (which he defines next) as an ordinary probability distribution. The event “ $\rho < .765$ ” can correctly be included in a sequence of events of the form “ $\rho < \rho_r$ ”, all having epistemic probability 1/20; but the other members of any such sequence (“almost surely”) don’t have  $\rho_r = .765$  (with some dubiously “random”  $\rho$ ’s floating around), but rather any sequence of fixed, predetermined, arbitrary, haphazardly taken, even aleatorily random values of  $\rho$ , random values of  $r$  whose distributions depend on them, and the corresponding values of  $\rho_r$  determined by  $r$ . (By contrast, when we wanted to define a “probability distribution for  $\rho$ ”, .765 would have to be a fixed value in repetitions of the same experiment - there is nothing of that here.)

I think one of the reasons why we still have problems with fiducial probabilities, is that we lack an adequate formalism for frequentist epistemic probabilities. For a start, let me offer the following. Given an ordinary, aleatory parametric model, consider a class of random “claims” or “statements”  $\{S\}$  depending on a random variable  $X$  on that model. In the simplest case, this is a single random statement (e.g., “ $\theta \geq X - 1$ ”), or a complementary pair of such statements. We can call a random claim “assessable” if it has the same (aleatory) probability under all parameter values; this probability value is taken to define the epistemic (since it is known to us) probability  $P$  of the random claim (cf. also [22]). We then define, for each  $\theta$ , a mapping  $V = V_\theta$  (depending on  $\theta$ ) from the space of possible realisations of the random claims to the two-point space  $\Omega = \Omega_S = \{\text{true}, \text{false}\}$  with  $V(S) = \text{“true”}$  if  $S$  is true. Obviously, the probability of a randomly (via  $X$ ) selected claim to be true is  $P$ , independently of  $\theta$  (but the set of all claims which are true is different for each  $\theta$ ). Hence we obtain a fixed probability distribution on  $\Omega$  (belonging to our random claim), independently of  $\theta$ .

If we have several random claims on the same probability space, we can build the cartesian product of the  $\Omega_S$ ’s, and no matter what happens with the joint distributions, the marginal distributions of the single  $\Omega_S$ ’s are still correct. Sometimes we may derive new assessable claims. In particular, if we have a sequence of independent claims (based on independent aleatory experiments), with the same probability of being true, we can apply the law of large numbers, and hence we can bet successfully (and even fairly)  $P : (1 - P)$  on the truth of any one claim (whose realisation is randomly selected by some  $X_S$ ) and will come out even in the long run. (The problem of selection among different possible random claims or successful bets is still not quite solved in the general theory (cf. [19]):

one idea has a deeper interpretation, and the other is mathematically more elegant.)

The “fiducial distribution of a parameter  $\theta$  for a given statistic  $T$ ” may be considered as a collection, or shorthand notation, for all (mutually compatible) successful claims or bets about  $\theta$  derivable from  $T$ . (Note that Fisher writes: “for a given statistic  $T$ ”!  $T$  will be different next time.) At most it might perhaps be called a “random distribution”, depending on  $T$ . From it can be derived epistemic probabilities and successful bets on events of the form “ $a < \theta < b$ ” (etc., by simple linear operations), but not, e.g., of the form “ $\theta^2 < a$ ” or “ $|\theta| < b$ ” (cf. [25]).

Replacing equal probabilities by an infimum of probabilities in the introduction of epistemic probabilities, most of the argument (except the fair bet aspect) can be done also with lower probabilities. There are some connections with belief function theory [27], but the interpretation is different.

It seems that Zabell ([31], e.g. p. 374), and probably many other readers, were confused by Fisher’s wild switching between aleatory and epistemic probabilities. But at least in his first papers, Fisher was, by intuition, always right; and the point (Fisher’s, subconsciously, and mine) is that we need both types of probability integrated in order to get a complete theory of statistics. One type is derived from the other in a perfectly objective way.

Returning to [7]: In his paper, Fisher finally compares fiducial and Bayes solutions. Curiously, he first argues by logic and not by insight that the two pertaining distributions “must differ not only numerically, but in their logical meaning” because the results will in general differ even though the Bayesian prior may be aleatorily true. But then he goes into details by considering a prior for which the posterior probability of  $\rho < .765$ , given  $r = .99$ , is not 5%, but 10%. He correctly argues that (with the Bayesian sampling) in 10% of all cases where  $r$  happens to be exactly  $= .99$ ,  $\rho$  will be less than  $.765$ . “Whereas apart from any sampling for  $\rho$  [!], we know that if we take a number of samples of 4, from the same or from different populations [!], and for each calculate the fiducial 5% value for  $\rho$ , then in 5% of all cases the true value of  $\rho$  will be less than the value we have found. There is thus no contradiction between the two statements. The fiducial probability is more general and, I think, more useful in practice ...” [exclamation marks added]. The sequences of events considered by both arguments in a sequence of experiments are clearly very different.

Here Fisher claims that if an (aleatory, true) Bayesian prior happens to be known, both methods give valid though different answers (for different questions). Later (e.g., in [11]), he strongly insists that the fiducial argument must only be used if nothing is known about the parameter. There is something to both attitudes. Clearly, the fiducial argument is correct and leads to successful bets even if a Bayesian prior is known. (By the way, this is also true if an inefficient statistic is used for the fiducial argument, a point against which Fisher later argues in his Author’s Note in [2], p. 428.) The problem is the efficiency, or the information, or the selection problem for successful bets alluded to above. According to present results, if an aleatory Bayesian method is available, it should be used - apart from questions of robustness or stability, the one big area Fisher refused to look at.

## 5 The aposteriori interpretation of confidence intervals

As all adherents of the Neyman-Pearson school know, a 95% confidence interval has a probability of 95% of covering the true unknown parameter apriori, before the data are in. After the data are in, the probability is 0 or 1, but we don’t know which one. That is all the theory says. But as probably most of those of us know who tried to teach Neyman-Pearson

statistics to critical, intelligent, unspoilt users of statistics, these scientists have a strong intuition that even “after the fact” there is, or should be, something with 95% probability; and they are very frustrated when they are told their intuition is entirely wrong. (Some may become overly critical of statistics as a whole, while some others will humbly believe the “experts”, like the werewolf in Christian Morgenstern’s German poem.)

Now the explanation of the conflict is simple. Both sides are right, in a way. Since Neyman considers only aleatory probabilities, for him 0 or 1 are the only possibilities. But the scientist using statistics can bet 19 : 1 that the unknown fixed parameter is in the fixed (but randomly derived) confidence interval, and in a long sequence of such bets with different independent experiments (and different confidence intervals with the same level), she will be right in 95% of all cases (or at least 95%, if she uses conservative confidence intervals), so her bets are successful. This means, she correctly has a frequentist epistemic probability (or lower probability, for the conservative intervals) of 95% for the event or claim that the parameter is covered, in full accordance with her intuition.

By the way, she would be rather stupid - though not wrong - offering the same bet again and again in the case of many independent replications of the SAME experiment, because after a while she could have learned much more about the parameter - unless the information that it was the same experiment was withheld from her, or some such artificial device.

It should be noted that the aposteriori interpretation of confidence intervals (and thus the implicit fiducial argument and a subconscious switch between aleatory and epistemic probability) was probably centuries old (cf. the related Endnote 8 in [31]); certainly around 1900 interpretations like “the odds are 1 : 1 that the true mean is within  $\pm 1$  probable error” were commonplace (cf., e.g., “Student’s” writings; cf. also the remarks on Maskell in [31], p. 371). It is Neyman’s merit that he clarified the purely aleatory argument; but by restricting himself to it, he cut himself off from one half of what statistics ought to be.

Incidentally, the above explanation can perhaps throw new light on the dispute between Neyman and Fisher around and about the invention of confidence intervals. At first, Fisher seemed to believe that Neyman’s intervals are essentially the same as his fiducial intervals (apart from the point of uniqueness and efficiency related to sufficiency etc.). But a short time later, he “mysteriously” seemed to change his mind and claimed that the two methods were very different, after all, without giving reasons. My guess and interpretation is that Fisher, more or less consciously, always included the epistemic interpretation with his intervals and in the beginning naively thought that Neyman did the same (given the formal similarity of what they wrote, and the lack of formalism for the epistemic aspect), until he suddenly (or perhaps creepingly) discovered that Neyman’s view was in fact much more primitive.

## 6 From lower probabilities to fiducial probabilities

As said before, with discrete models we have to use lower probabilities to describe successful bets. Moreover, even if ideally we have a model with a fiducial proper probability, in reality (e.g., with a digital computer) the data are always discretized. But when the discretization gets finer and finer, the lower probabilities of an event and its complement converge to proper probabilities adding to one.

A simple example is the following [14]. Let  $X$  be uniformly distributed on  $[\theta, \theta + 1]$  ( $\theta$  real), and let (for every  $n$ )  $Y_n$  be  $X$  rounded upwards to the nearest multiple of  $1/n$ . Then for every  $c$  between 0 and 1, and all  $\theta$ ,  $P_\theta(Y_n \leq \theta + c) \leq c$ , and  $P_\theta(Y_n > \theta + c) \leq 1 - c + 1/n$ , hence we can bet with epistemic lower probability  $1 - c$  on  $[\theta < y_n - c]$ , and with odds

$(c - 1/n) : (1 - c + 1/n)$  on  $[\theta \geq y_n - c]$ . The sum of the lower probabilities is  $1 - 1/n \rightarrow 1$  as  $n \rightarrow \infty$ .

Thus, fiducial probabilities are just a limiting case of lower probabilities, though interesting in their own right because they allow fair (pairs of) bets (with a frequentist interpretation!). Hence they produce something similar to the Bayesian omelette, after all, without breaking the Bayesian eggs.

## References

- [1] M. S. Bartlett. Complete simultaneous fiducial distributions. *Ann. Math. Statist.* 10:129–138, 1939.
- [2] J. H. Bennett., Ed. *Collected Papers of R. A. Fisher*, vol. II (1925-31). The University of Adelaide, 1972.
- [3] D. R. Cox, and D. V. Hinkley. *Theoretical Statistics*. Chapman and Hall, London, 1974.
- [4] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* 38:325–339, 1967.
- [5] A. P. Dempster. A generalization of Bayesian inference. *J. Roy. Statist. Soc. B* 30:205–245, 1968.
- [6] B. Efron. R. A. Fisher in the 21st Century: Invited paper presented at the 1996 R. A. Fisher Lecture. *Statistical Science* 13, 2:95–122, 1998.
- [7] R. A. Fisher. Inverse probability. *Proc. of the Cambridge Philosophical Society* 26:528–535, 1930. Reprinted in *Collected Papers of R. A. Fisher*, ed. J. H. Bennett, Volume 2, 428–436, University of Adelaide 1972.
- [8] R. A. Fisher. The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proc. Roy. Soc. London, Ser. A* 139:343–348, 1933.
- [9] R. A. Fisher. The fiducial argument in statistical inference. *Annals of Eugenics* 6:391–398, 1935.
- [10] R. A. Fisher. Statistical methods and scientific induction. *J. Roy. Statist. Soc. Ser. B* 17:69–78, 1955.
- [11] R. A. Fisher. *Statistical Methods and Scientific Inference*. Oliver and Boyd, London, 1956. (2nd ed. 1959, reprinted 1967).
- [12] D. A. S. Fraser. *The Structure of Inference*. Series in Probability and Mathematical Statistics. Wiley, New York, 1968.
- [13] F. Hampel. Fair bets, successful bets, and the foundations of statistics. Abstract, Second World Congress of the Bernoulli Society, Uppsala, 1990.
- [14] F. Hampel. Some thoughts about the foundations of statistics. In *New Directions in Statistical Data Analysis and Robustness*, S. Morgenthaler, E. Ronchetti, and W. A. Stahel, Eds. Birkhäuser Verlag, Basel, 125–137, 1993.

- [15] F. Hampel. On the philosophical foundations of statistics: Bridges to Huber's work, and recent results. In *Robust Statistics, Data Analysis, and Computer Intensive Methods; In Honor of Peter Huber's 60th Birthday*, H. Rieder, Ed., no. 109 in Lecture Notes in Statistics, Springer-Verlag, New York, 185–196, 1996.
- [16] F. Hampel. What can the foundations discussion contribute to data analysis? And what may be some of the future directions in robust methods and data analysis? *J. Statist. Planning Infer.* 57:7–19, 1997.
- [17] F. Hampel. How different are frequentist and Bayes solutions near total ignorance? In *Proc. 51th Session of the ISI, Contrib. Papers, Book 2, Istanbul* 25–26, 1997.
- [18] F. Hampel. On the foundations of statistics: A frequentist approach. In *Estatística: a diversidade na unidade*, M. S. de Miranda and I. Pereira, Eds., Edições Salamandra, Lda., Lisboa, Portugal, 77–97, 1998.
- [19] F. Hampel. An outline of a unifying statistical theory. In *Proc. of the 2nd Internat. Symp. on Imprecise Probabilities and their Applications*, Cornell University, 26–29 June 2001, G. de Cooman, T. L. Fine, and T. Seidenfeld, Eds. Shaper Publishing, Maastricht, 205–212, 2000.
- [20] F. Hampel. Why is statistics still not unified? In *Proc. 53rd Session of the ISI, Contrib. Papers, Book 1, Seoul* 73–74, 2001.
- [21] A. Kolmogoroff. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Julius Springer, Berlin, 1933.
- [22] A. N. Kolmogorov. The estimation of the mean and precision of a finite sample of observations. In Russian. (Section 5: Fisher's fiducial limits and fiducial probability). *Bull. Acad. Sci. U.S.S.R., Math. Ser.* 6:3–32, 1942.
- [23] J. Neyman. 'Inductive behaviour' as a basic concept of philosophy of science. *Rev. Internat. Statist. Inst.*, 25:7–22, 1957.
- [24] J. Neyman. Foundations of behavioristic statistics. In *Foundations of Statistical Inference*, V. P. Godambe, and D. A. Sprott, Eds. Holt, Rinehart and Winston of Canada, Toronto, 1971.
- [25] P. J. G. Pitman. Statistics and science. *J. Amer. Statist. Assoc.* 52:322–330, 1957.
- [26] C. R. Rao. *Linear Statistical Inference and Its Applications*, Wiley, London, 1965. (2nd edition 1973).
- [27] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N. J., 1976.
- [28] J. W. Tukey. A smooth invertibility theorem. *Ann. Math. Stat.* 29:581–584, 1958.
- [29] J. W. Tukey. Handouts for the Wald Lectures 1958. In *The Collected Works of John W. Tukey, Volume VI: More Mathematical (1938 – 1984)*, C. L. Mallows, Ed. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA. Ch. 10, 119 – 148, 1990.
- [30] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

- [31] S. L. Zabell. R. A. Fisher and the fiducial argument. *Statistical Science* 7, 3:369–387, 1992.