An outline of a unifying statistical theory

Frank Hampel
Seminar for Statistics of ETH Zurich
CH-8092 Zurich, Switzerland

Abstract

A new statistical theory is outlined which builds a bridge between frequentist and Bayesian approaches and very naturally uses upper and lower probabilities. It started with an attempt to investigate how far one can get with a frequentist approach; this approach goes beyond the Neyman-Pearson and the Fisherian theory in explicitly using intersubjective epistemic upper and lower probabilities allowing an operational frequentist interpretation (not tied to repetitions of an experiment), and in deriving what is valid of Fisher's mostly misinterpreted fiducial probabilities as a very special case within a broader framework. It formally contains the Bayes theory as an extremal special case, but at the other extreme it also allows starting with the state of total ignorance about the parameter in an objective, frequentist learning process converging to the true model, thereby solving a problem of artificial intelligence (AI). The general theory describes (rather similar) optimal compromises between frequentist and Bayesian approaches within (and outside) either framework, thus also providing a new class of "least informative priors". There is also a connection with information theory. Key concepts are "successful bets", more specifically "least unfair successful bets", "cautious surprises", and "enforced fair bets", including "best enforced fair bets". The main emphasis is on prediction. When going from inference to decisions, upper and lower probabilities (which avoid sure loss) are replaced by proper probabilities (which are coherent), somewhat analogous to Smets' pignistic transformation of belief functions. Much still needs to be done, but several examples for the binomial (the "fundamental problem of practical statistics") have been worked out, and there are also first (rather limited) solutions for continuous one-parameter situations, including their robustness problem.

Keywords. Foundations of statistics, frequentist approach, Bayesian approach, Neyman-Pearson theory, Fisherian theory; upper and lower probabilities, aleatory and epistemic probabilities, frequentist intersubjective epistemic probabilities, fiducial probabilities; successful bets, least unfair successful bets, best enforced fair bets; cautious surprises; state of total ignorance, objective learning process; inference and decisions; prediction for binomial data.

1 Introductory notes

1.1 Some general remarks

The work outlined here grow out of an attempt to clarify the foundations of traditional statistics (including Bayesian statistics, but not including belief function theory and other more recent developments, though there turned out a number of formal parallels with them). The results went far beyond traditional statistics and contained a number of surprises for me.

For one thing, I became convinced that upper and lower probabilities should be a central tool in statistics, rather than an exotic marginal structure.

Secondly, I now believe that the distinction between aleatory probabilities (which exist, ore are tentatively supposed to exist, "objectively" in Nature, usually unknown to us) and epistemic probabilities (referring to what we actually know, or believe to know) is important to explain and dissolve the discrepancies between different schools of statistics. Neyman explicitly considered only aleatory probabilities, while both "objective" and "subjective" Bayesians consider only epistemic probabilities, hence the twain can never meet (at least strictly speaking). Fisher worked mostly with frequentist aleatory probabilities (like Neyman), but I now consider his fiducial argument the first attempt (wrong in execution, but correct in principle and vision) to introduce frequentist epistemic probabilities on a broader scale and thus bridge the gap between aleatory and epistemic views.

The fiducial argument intrigued, puzzled and inspired many great statisticians, including Kolmogorov, Tukey, Dempster, Seidenfeld, Fine, and many others; but to my limited knowledge, probably all this work, with the remarkable exception of [30], was based on Fisher's own false interpretation leading to a probability distribution on the parameter space. Perhaps my greatest surprise was that the — proper — fiducial argument came out naturally as a very special case of my theory, thus embedding it into a much broader framework, including also discrete distributions (for the price of upper and lower probabilities). In this framework, also a (nontrivial) aposteriori interpretation of confidence intervals (by frequentist intersubjective epistemic lower probabilities) becomes possible (impossible and strictly forbidden in the Neyman-Pearson theory, but rightly requested by all "unspoiled" intelligent users of statistics); and this fact may shed some new light on the early debate between Fisher and Neyman on confidence intervals.

A final surprise was a certain symmetry between the Bayesian approach and my extended frequentist approach, leading to solutions in either framework closest to the other one, or even in between both frameworks. For various reasons, there is a stress on prediction, rather than parameter estimation. Throughout the work, I make free use of the age-old equivalence between probabilities and bets: it turns out that two-sided (in particular Bayesian "fair") bets correspond to proper probabilities, while one-sided bets correspond to lower (and hence upper) probabilities.

1.2 A prototype example for the new theory

Consider a sequence of "independent, identical" experiments with two possible outcomes ("success" or "failure") in each case (i.e., a sequence of independent Bernoulli trials), and assume that nothing is known apriori about the probability of success θ. Given that in n (≥ 0) past trials, X successes have been observed, how can one bet on the event that in k (≥ 1) future trials, the number of successes Y lies in some given set A? (E.g., a new medical treatment, with no previous experience, has been successful in 3 out of 5 cases; what are the "chances" that it will be successful in exactly 2, or at least 2, cases out of 4 future cases?) Cf. also [28].

Let $\underline{m}$ $(0 \leq \underline{m} \leq 1)$ denote a kind of lower probability leading to the (one-sided!) bet $\underline{m} : (1 - \underline{m})$ on the event considered. A natural desideratum is that the *expected gain* EG of the bet is ≥ 0. Clearly, this cannot be achieved in general for all θ conditionally on any given $X = x$ (except by the trivial bet $0 : 1$), because x can always be "misleading" by chance. But, amazingly, it can be achieved if we are allowed to average also over the X's that "might have been observed", that is, if we take the joint expectation over the distribution of X and Y. By this step, we lose full conditionality, to be sure, but we gain a frequentist epistemic statement about the "chances" of A (while Bayes solutions are conditional, but in general have no objective frequentist interpretation). We call such a class of bets on A (one for each x) "successful" for short (instead of : "not unsuccessful on the average"). More formally:

Definition 1 *(Starting with total ignorance about θ) a (class of) bet(s) on A (one for each x), described by $\underline{m}(Y \in A|.)$, is called successful iff*

$$E_\theta G = E_\theta^X E_\theta^Y G \geq 0 \quad \forall \theta,$$

or equivalently iff

$$E_\theta^X \underline{m}(Y \in A|X) \leq P_\theta(Y \in A) \quad \forall \theta.$$

(The superscript denotes the random variable with respect to whose distribution the expectation is taken.)

Example: Let $n = k = 1$, $A = \{1\}$.
If $X = 0$, we must bet $0 : 1$, because θ might be 0. By contrast, if $X = 1$, any bet $c : (1 - c)$ $(0 \leq c \leq 1)$, combined with the trivial bet for $X = 0$, is successful, because $E_\theta G = P_\theta(X = 1) \cdot ((1 - c) \cdot P_\theta(Y = 1) - c \cdot P_\theta(Y = 0)) + P_\theta(X = 0) \cdot (1 \cdot P_\theta(Y = 1) - 0 \cdot P_\theta(Y = 0)) = \theta \cdot ((1-c)\theta - c(1-\theta)) + (1-\theta) \cdot \theta = \theta \cdot (1-c) \geq 0 \; \forall \theta$.

As in this example, in general an infinity of successful bets exists, and the question arises which one to select. A naive idea would be to select the (or an) "admissible" (i.e. the — or a — most extreme) one, in our example $1 : 0$ if $X = 1$ (and of course $0 : 1$ if $X = 0$). But there is a deeper idea.

Bayesians rightly argue that if one has to make a decision (as opposed to inference, which for them is effectively the same), one has to act as if one has a probability distribution. Correspondingly, the Bayesians consider (only) two-sided bets, where the opponent can change sides. If he/she can change sides freely depending on each x observed, in general successful bets are not possible anymore; but we can ask for the (class of) "fair" bet(s) that minimizes the maximum expected loss under all changes of sides by the opponent. They are called "best enforced fair bets" and are in a sense the Bayesian solutions (with corresponding distinguished priors) coming closest to fulfilling a frequentist requirement. (In our example, with the natural linear loss function, these bets are: $3 : 1$ on the same event again, and $1 : 3$ on the opposite event, with minimax risk $R_e := 1/4$.)

Returning to the selection problem for successful bets, we can symmetrically also ask for the successful bets which come closest to a Bayesian solution (viz., which minimize the maximum risk under all changes of sides among all **successful** bets). They are called "least unfair successful bets".

In our example, when we offer a successful bet with $c : (1 - c)$ if $X = 1$, assume our opponent knows θ (this is allowed). He/she will always switch sides if $X = 0$, making our conditional (on $X = 0$) expected gain $= -\theta$. If $X = 1$, he/she will switch iff $\theta \underset{(=)}{>} c$, making our conditional expected gain $= -|\theta - c|$. Our overall expected gain is then $-\theta(1-\theta) - |\theta - c|\theta$. The c which minimizes the maximum expected loss over all θ is given by $1 - c = 6 - 4\sqrt{2}$, hence $c \approx 0.6569$ (closer to 1 than to 0, as seems intuitively reasonable). The minimax risk is $R_f := 6 - 4\sqrt{2} \approx 0.3431$.

It is to be expected that the frequentist and Bayesian statistics closest to each other become quickly very similar as more and more information is contained in the past, that is, as n grows.

In general, we can compute for every bet its maximum risk r_1 when considered as one-sided bet (≤ 0 for successful bets, ≥ 0 for Bayesian fair bets), and its maximum risk r_2 when considered as two-sided bet ($\geq R_f \geq R_e$ for successful bets, $\geq R_e$ for Bayesian bets). We thus have reached a certain symmetry between frequentist and Bayesian solutions and can consider in the "risk set" of the pairs (r_1, r_2) in $\mathbb{R}^2$ also the "admissible" lower left boundary with procedures which are neither quite frequentist nor quite Bayesian, but may be useful compromises between the two classes.

2 Overview over the theory

2.1 Introduction

Since it is not possible to describe the details of the new theory in a short talk or a short paper, I shall try to give a rather nontechnical overview with an introductory guide to the more detailed literature and an outline of the past achievements and problems and possible future developments as I see them presently.

It should be noted that parts, pieces and fragments of the theory did already exist. However, the way of combining them, enlarging them, and filling the gaps by means of new concepts appears to be new - somewhat surprisingly. In particular, I was amazed to find out during my research how incomplete the existing frequentist theories of statistics were.

The new theory uses and investigates (upper and lower) probabilities and not, for example, belief functions (cf., e.g., [6], [7], [31], [33]), possibility functions [41], [8], or other "fuzzy" concepts [40], which have their own realms of application; but it has in common with the former theories that it uses pairs of numbers to describe incomplete knowledge and thus can also model the state of total ignorance in an appropriate way. It differs from them by using different rules, and by the numbers not being subjective, but having an objective and operational frequentist interpretation.

As indicated in the abstract, the framework of the new theory contains the Neyman-Pearson, Bayesian (with "robust Bayesian") and Fisherian theory, including what is correct about fiducial probabilities (see below); it makes strong use of likelihoods, solves a major problem of artificial intelligence, and also has a close connection with information theory.

The theory extensively utilizes the old relation between probabilities and bets, or odds ratios. Proper probabilities are equivalent to two-sided bets (or pairs of bets, on an event A and its complement), and in the same way upper and lower probabilities can be seen as equivalent to one-sided bets. (For one-sided bets, cf. also [35].) While Bayesians like to consider fair (pairs of) bets, with (mostly only subjectively) expected gain equal zero, the theory also studies (one-sided) "successful bets," with (objectively) expected gain greater or equal zero [18], [25].

A closely related concept is that of "cautious surprises" [18], which is tied to information theory. It is partly stronger, but not as nicely linear as successful bets, and therefore its exploration has been largely postponed.

There is also a partly new interpretation of the fundamental paper by Bayes [1], who was basically a frequentist, in the light of the new theory ([14]; cf., e.g., [25]).

2.2 The basic framework for inference

The main framework is that of a given parametric model (e.g., independent Bernoulli trials), with possibly some prior knowledge about the parameter, with a past observation $X = x$, a future observation Y and a given event A in the range of Y about whose occurrence some claims shall be made. The stress on prediction is not accidental (cf. [22], [25]). A may also depend on X, as in usual prediction intervals. If, in a variant of the basic framework, A is a subset of the parameter space, it has to depend on X (as with confidence intervals and fiducial probabilities), otherwise no nontrivial inference is possible in this theory.

Since our one-sided bet that Y will be in A, may depend on the past observation x, we actually have to consider a whole class of conditional bets, given x. If we wanted to bet "successfully," that is, with expected nonnegative gain, conditionally given any x, we could not learn from the distribution of X, since in general x may always be an extremely unlikely and extremely misleading observation, and we would have to take the worst possible case into account. Hence, in order to obtain the frequentist property of being successful, we have to be able to compensate for very unlikely x's, and this can most simply be done by averaging the expected gain over the full distribution of X. Such a (class of) bet(s) will be called "successful" [18], [25], and amazingly enough, this definition works. However, we have to give up full conditionality and hence coherence [37]; but we still avoid sure loss. If we try to interpret this in monetary terms, it means we might perhaps have gained a bit more money, but at least we did not lose any money on the average (as Bayesians, despite their claims, very often do, cf. [25]). A further simple thought shows that we can always modify our bets so that we never lose money for sure even in single cases. - For further aspects of the one- and two-sided betting situation, see [25].

2.3 Fiducial probabilities

A surprising side result was the clarification of what is correct about the fiducial argument ([9], [10]; for a nice historical survey by a nonspecialist, see [39]). Fiducial probabilities have been grossly misunderstood by almost everybody, including Fisher himself. A remarkable exception is the paper by Pitman[30] (I owe this reference to R. Staudte) who describes in a clean and mathematically oriented way what the (proper) fiducial argument actually achieves. Contrary to common belief (which has led to the well-known counterexamples), it does not lead to a probability distribution on the parameter space, but to (frequentist epistemic) probabilities for the correctness of statements of the kind "$\theta < X + c$" (e.g. "$\theta < 2 + 3''$ for $c = 3$ if $X = 2$) which are random statements depending on the random X (while θ is still a fixed unknown constant). These (proper) probabilities describe successful bets which are even fair in the Bayesian sense (thus making something similar to the Bayesian omelet after all, without breaking the Bayesian eggs), but their framework and interpretation is different from the Bayesian one. For example, in the simplest case ($X \sim N(\theta, 1)$), the fiducial argument and an improper Bayesian prior lead to formally exactly the same formula, but the proper interpretation, including the underlying probability space (which is suppressed and hidden by the usual notation), is entirely different.

The (proper) fiducial argument is also closely related to the appropriate frequentist aposteriori interpretation of confidence intervals (cf. above) which leads to a series of independent bets, one for each of a series of (normally different!) independent experiments, with the appropriate (minimum) long run rate of successes. Fiducial probabilities are derived as a special case of a sideline (claims for random parameter sets, instead of prediction of future events) of the new theory; they are thus put into a much broader framework, which also bridges the gap between continuous and discrete distributions [18].

2.4 The decision problem

So far we have considered the inference problem of how to describe in a quantitative and operationally verifiable way our incomplete knowledge after obtaining an observation $X = x$ from a given parametric model. The Bayesian claim that only proper probabilities must be used ("enforced fair bets") is not true for inference (cf. also [34]), but appears to be true for decisions. Thus, in vague analogy to Smets' "pignistic transformation" [32], the theory replaces, for decision purposes, the upper and lower probabilities by proper probabilities which cannot be successful anymore (a few special cases excepted), but which minimize the maximum expected loss ("best enforced fair bets and probability distributions," cf. [19], [25]). They can usually (but not always) be obtained as special Bayes solutions, from what may be called a new class of "least informative priors." These Bayes solutions (the "least unsuccessful Bayes solutions") are closest to being frequentist solutions in a minimax sense, and the remaining gap can be measured by the minimax risk.

2.5 A bridge between frequentist and Bayesian statistics

The solution for two-sided bets suggests a solution for the selection of a particular successful bet in the one-sided betting situation: choose the successful bet which minimizes the maximum risk in the two-sided betting situation ("least unfair successful bets"). Naturally, this minimax risk is at least as big as the one without restriction to successful bets. The corresponding bet is the one with frequentist interpretation which in a sense comes closest to being also a Bayesian solution, with the pertaining properties.

A few examples for the binomial situation (the "fundamental problem of practical statistics," cf. [28], [29]) have already been worked out [36], [23], [25]. They look reasonable and show that the best compromise frequentist and Bayesian solutions are not too far apart. More generally, we may assign to any one-sided bet its maximum one-sided and two-sided risk and look at the set of all attainable risk vectors in two dimensions, as in decision theory; this set contains the two compromise solutions mentioned above at the extremes of the interesting "admissible" lower left "edge" of the risk set, and in addition presumably many "admissible" solutions in between, which are neither quite frequentist nor quite Bayesian, but which build a natural bridge between the two approaches. In particular, ϵ-successful and ϵ-Bayes solutions might conceivably contain a marked overall improvement over the "pure" classes.

2.6 Some other aspects

Besides the binomial situation, there are also some first results for the Poisson distribution (M. Wolbers, orally) and for rather general continuous one-parameter problems, including a very first discussion of the problems arising from the fact that parametric models are almost never exact (the binomial often being an exception with a high degree of accuracy), namely the "robustness problem" ([21]; for the background see [13], Ch. 1 and Ch. 8.1, and [27]).

The theory provides a new and sometimes surprising outlook on customary problems in statistics. Thus, in the testing situation, one (obviously) cannot bet successfully (and nontrivially) on a fixed hypothesis; however, one can bet successfully on the correctness of a test decision [18].

For introductory reading on the theory, the three main papers are [18], [21] and [25]; compare also [20] and[22] for a wider horizon. [18] is the broadest paper on the theory, but naturally least advanced; [21] provides some further solutions which have not yet been followed up; and [25], probably the most readable of the three, pushes farthest what so far has been the main line of development.

3 Remarks on the previous papers about the theory

3.1 The beginning

The following sections contain some remarks on the development of the theory and its open problems, as an aid to a deeper and better understanding of the pertaining literature, and possibly as a stimulus for further research.

The first tentative concept in the emerging theory was the "Moeglichkeit" ([14], reproduced in [16]; [18]), not to be confused with the "possibility" by Zadeh [41] and Dubois and Prade [8]. It has some esthetic reasons in its favor, but after [18] it disappeared into the background. However, I think it might become useful as a starting point once the theory for cautious surprises is being worked out.

3.2 Successful bets and related concepts

The next concept was the key concept of successful bets ([15], cf. also [17], both reproduced in [16]). It led to a rich and multifaceted paper [18] containing a number of germs for future research. In retrospect, the following points may be noted:

The unifying formalism (loc. cit., p. 127) contains both the start with total ignorance about the parameter (given the model), and the start with a Bayesian prior, as extreme cases. The bridge in between is still little explored (and still not fully formalized, although a Choquet [4] capacity of order 2 on the parameter space might be quite suitable as upper bound for the priors). The bridge contains as examples restrictions on the parameter space, and certain neighborhoods of a Bayesian prior (cf. [26]; [27], p. 263f; and Berger's [2] work on the so-called "robust Bayesian" approach, which is actually only "half-robust" since it does not consider neighborhoods of the parametric model).

A little note on notation ([18], p. 127) in answer to a frequent question: the reason for writing the upper and not the lower probability as m without a bar, and using it preferentially, was a far-reaching guess that eventually cautious surprises will become the central concept, although for successful bets the lower probabilities would be more convenient.

The following sections (loc. cit.) on successful bets and on fiducial probabilities (cf. also [15], [16], [17]) contain much material which apparently has to be repeated again and again.

The "surprise" ([18], p. 131, cf. also [16], [17]), closely related to entropy, is not the only one in the literature. For example, Good ([11]; the version in [12] is incomplete), in the spirit of a pure mathematician, defines a whole class of "surprises" whose (according to Good, orally) most important member differs from the above concept just by an additive constant.

The theorem on the relation between cautious surprises and successful bets ([18], p. 131) is obviously derived under a start with total ignorance (as H. Carnal - orally - noticed in 1995), but this assumption was forgotten to be stated explicitly.

The various likelihood-based approaches (loc. cit., p. 132f) still wait for further exploration; but the proposal and form of solution in the last paragraph - minimax MSE and maximum likelihood plus a constant - have recently gained increased interest as possible alternatives to least unfair successful bets, due to their simple structures (needed in more general situations), the results for the normal and other distributions [21], and a remark by M. Wolbers (orally, 1999) on the relation to unbiased estimation (cf. also the remarks on best enforced fair bets below).

For the examples with $n = k = 2$ ([18], p. 134), Steiner [36] found also admissible asymmetric solutions, so that the claim about symmetrizing is wrong. However, it still seems reasonable to generally restrict oneself to symmetric solutions.

3.3 Enforced fair bets

The next major research step was the definition of "best enforced fair bets," a solution for the decision problem, in [19]. In this short and highly condensed paper, the formula is given for the solution with the minimax expected monetary loss if the opponent is allowed to choose sides depending on the past observation x, a strong generalization of simple fair bets to the case of a set of conditional bets, given x. (Cf. the detailed derivation in [25].) Unfortunately, the most clever change of sides leads to the absolute value instead of, for example, a square which would be easier to handle mathematically (but then we would need "squared" money); compare also the remark on MSE above.

The examples given (and more fully described in [25]) are of general conceptual interest. They show that in a very famous philosophical problem (how to bet in the state of total ignorance), in the simplest case the "principle of insufficient reason," symmetric Bayesian priors, Smets' pignistic transformation [32], and the best enforced fair bets yield the same solution, though with a different derivation each time; and in other cases, all solutions differ; moreover, the "least informative priors" for the "least unsuccessful Bayes solutions" (cf. above) differ depending on the problem considered, a very anti-Bayesian situation found, remarkably, also in Bernardo's [3] "reference priors," but probably nowhere else in Bayesian theory. In addition, some (unimportant?) "best enforced probability distributions" cannot be obtained as Bayes solutions.

In another simple example, it is shown that Laplace's "rule of succession," which found a natural interpretation in terms of the "Moeglichkeit" [14], is not optimal in the present context.

3.4 Successful bets in different parametric models

The second somewhat larger paper [21] mainly extends the range of applicability of successful bets. It also contains a sketch of a proof for asymptotic consistency. Successful bets are derived for the one-parameter normal and the ("nonregular") exponential with shift parameter, by "transforming" likelihoods into upper and lower probabilities, and asymptotically for general "regular" one-parameter models. The sketch of a coarse first treatment of the robustness problem makes strong use of the heuristics of approximate parametric models (cf., e.g., [13]) and uses a nonstandard type of asymptotics (cf. [24] for some general critical comments on asymptotics).

3.5 Least unfair successful bets, and an outlook

The last major research step was the selection of a canonical successful bet out of the infinity of successful bets, namely the one that is closest to being Bayesian ("least unfair successful bets"), which allows also the definition of a class of optimal compromise solutions between frequentist and Bayesian ones ([25], cf. also [36] and [23]). For this solution of the uniqueness problem for successful bets, the "detour" via enforced fair bets was needed. This "linear" part of the overall theory (cf. [18] for a broader perspective) is now well rounded off and rather encompassing, containing and enlarging practically all the customary statistical theories (including a critical appraisal of subjectivist Bayesianism from a higher perspective), and several examples for the fundamental case of 2 binomials have been worked out by Steiner [36] and the author, showing that reasonable looking solutions exist and can be found.

However, this view is a bit too rosy. The numerical solutions are too complicated for general use; most had to be found by brute force on the computer. I think either they shall serve as benchmarks for a simple approximation which can be generally used in practice; or some concepts shall be modified to yield a more tractable mathematics (or both simultaneously). In particular, it is tempting to replace the absolute value signs in several formulas by squares; even Bayesians might prefer the mean of the posterior to the median, although the monetary interpretation does not seem clear at present.

There is also a question whether the requirement of successful bets may be in a sense "too strong," since some rather nice looking proposals for inference, for example by Walley [38] and especially by Coolen [5] do not fulfill it. This may have to do with the very strong property that successful bets lead out of the state of total ignorance without any additional arbitrariness. Perhaps weakening the criterion by an epsilon might have a big effect, but this would still leave the present theory as a basis for comparison.

One weak point of the theory seems to be both unavoidable and minor: it would be nice to have a method which is both fully coherent and fully successful. However, within my framework this appears to be impossible. If the probability of an event can be arbitrarily close to zero under different parameters, I do not see how one can bet on it successfully and nontrivially without "borrowing" from somewhere else. On the other hand, the lack of coherence (or else of successfulness, as with optimal Bayesian compromises) can be quantitatively assessed, is rather limited and disappears asymptotically, as more and more information comes in.

References

[1] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philos. Trans. Roy. Soc. London*, A 53:370–418, 1763. Reprinted in *Biometrika* **45** (1958), 293–315, and in Pearson, E. S., Kendall, M. G. (eds.)(1970): *Studies in the History of Statistics and Probability,* 131–153, Griffin, London.

[2] J. O. Berger. The robust Bayesian viewpoint. In J. B. Kadane, editor, *Robustness of Bayesian Analyses*. Elsevier Science, Amsterdam, 1984.

[3] J. M. Bernardo. Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc.* Ser. B, 41:113–147, 1979.

[4] G. Choquet. Theory of capacities. *Ann. Inst. Fourier*, 5:131–295, 1953/54.

[5] Frank P. A. Coolen. Low structure imprecise predictive inference for Bayes' problem. *Statistics and Probability Letters*, 36:349–357, 1998.

[6] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38:325–339, 1967.

[7] A. P. Dempster. A generalization of Bayesian inference. *J. Roy. Statist. Soc.*, B 30:205–245, 1968.

[8] D. Dubois and H. Prade. *Theory of Possibility.* Plenum, London, UK., 1988. Original Edition in French (1985) Masson, Paris.

[9] Ronald A. Fisher. Inverse probability. *Proc. Camb. Phil. Soc.*, 26:528–535, 1930. Reprinted in Collected Papers of R. A. Fisher, ed. J. H. Bennett, Volume 2, 428–436, University of Adelaide 1972.

[10] Ronald A. Fisher. *Statistical Methods and Scientific Inference*. Oliver and Boyd, London, 1956. (2nd ed. 1959, reprinted 1967).

[11] I. J. Good. The probabilistic explication of information, evidence, surprise, causality, explanation, and utility. In V. P. Godambe and D. A. Sprott, editors, *Foundations of Statistical Inference*, pages 108–141. Holt, Rinehart, and Winston, Toronto, 1971. Reprinted in Good (1983).

[12] I. J. Good. *Good Thinking; The Foundations of Probability and Its Applications.* University of Minnesota Press, Minneapolis, 1983.

[13] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions.* Wiley, N. Y., 1986.

[14] Frank Hampel. Zu den Grundlagen der Statistik. Research Report 58, Seminar für Statistik, ETH Zurich, 1989. English translation: On the foundations of statistics. Proc. 47th Session ISI, Paris 1989, Contrib. Papers, Book 1, 423–424.

[15] Frank Hampel. Fair bets, successful bets, and the foundations of statistics. Abstract, Second World Congress of the Bernoulli Society, Uppsala, 1990.

[16] Frank Hampel. Some remarks on the foundations of statistics. Research Report 66, Seminar für Statistik, ETH Zurich, 1991. pp. 14.

[17] Frank Hampel. Some remarks on the foundations of statistics. In *Bull. of the ISI, Proc. 48th Session, Volume LIV, Book 4(B), Cairo*, pages 553–554, 1991.

[18] Frank Hampel. Some thoughts about the foundations of statistics. In S. Morgenthaler, E. Ronchetti, and W. A. Stahel, editors, *New Directions in Statistical Data Analysis and Robustness*, pages 125–137. Birkhäuser Verlag, Basel, 1993.

[19] Frank Hampel. Predictive inference and decisions: Successful bets and enforced fair bets. In *Proc. 49th Session of the ISI, Contrib. Papers, Book 1. Firenze (Italy)*, pages 541–542, 1993.

[20] Frank Hampel. Zur Grundlagendiskussion in der Statistik. In J. Frohn, U. Gather, W. Stute, and H. Thöni, editors, *Applied statistics - recent developments; Pfingsttagung 1994 der Deutschen Statistischen Gesellschaft, Festkolloquium zur 20-Jahrfeier des Fachbereichs Statistik, Universität Dortmund*, Sonderhefte zum allgemeinen statistischen Archiv, Heft 29, pages 131–148. Vandenhoeck & Ruprecht, Göttingen, 1995.

[21] Frank Hampel. On the philosophical foundations of statistics: Bridges to Huber's work, and recent results. In Helmut Rieder, editor, *Robust*

Statistics, Data Analysis, and Computer Intensive Methods; In Honor of Peter Huber's 60th Birthday, number 109 in Lecture Notes in Statistics, pages 185–196. Springer-Verlag, New York, 1996.

[22] Frank 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.

[23] Frank Hampel. How different are frequentist and Bayes solutions near total ignorance? In *Proc. 51th Session of the ISI, Contrib. Papers, Book 2, Istanbul*, pages 25–26, 1997.

[24] Frank Hampel. Is statistics too difficult? *Canad. J. Statist.*, 26(3):497–513, 1998.

[25] Frank Hampel. On the foundations of statistics: A frequentist approach. In Manuela Souto de Miranda and Isabel Pereira, editors, *Estatística: a diversidade na unidade*, pages 77–97. Edições Salamandra, Lda., Lisboa, Portugal, 1998.

[26] P. J. Huber. The use of Choquet capacities in statistics. In *Proc. 39th Session of the ISI*, volume 45, pages 181–188 (discussion: 189–191), 1973.

[27] P. J. Huber. *Robust Statistics.* Wiley, N. Y., 1981.

[28] K. Pearson. The fundamental problem of practical statistics. *Biometrika*, 13:1–16, 1920.

[29] K. Pearson. Note on the 'fundamental problem of practical statistics'. *Biometrika*, 13:300–301, 1921.

[30] E. J. G. Pitman. Statistics and science. *J. Amer. Statist. Assoc.*, 52:322–330, 1957.

[31] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, N. J., 1976.

[32] Philippe Smets. Constructing the pignistic probability function in a context of uncertainty. In M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, volume 5, pages 29–39. Elsevier Science Publ., 1990.

[33] Philippe Smets. The transferable belief model and other interpretations of Dempster-Shafer's model. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, volume 6, pages 375–383. Elsevier Science Publ., 1991.

[34] Philippe Smets. No Dutch Book can be built against the TBM even though update is not obtained by Bayes rule of conditioning. Technical Report TR/IRIDIA/93–9, Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle, Bruxelles, 1993.

[35] C. A. B. Smith. Consistency in statistical inference and decision. *J.Roy. Statist. Soc.*, B 23:1–37, (with discussion), 1961.

[36] Rolf Steiner. Erfolgreiche Wetten für die Binomialverteilung. Diplomarbeit, Seminar für Statistik, Swiss Federal Institute of Technology (ETH) Zurich, 1995.

[37] Peter Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

[38] Peter Walley. Inferences from multinomial data: Learning about a bag of marbles. *J. R. Statist. Soc. B*, 58(1):3–57, 1996. With discussion.

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

[40] L. A. Zadeh. Fuzzy sets. *Inform. Control*, 8:338–353, 1965.

[41] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.