A sketch of a unifying statistical theory

A SKETCH OF A UNIFYING STATISTICAL THEORY

by

FRANK HAMPEL

Research Report No. 86
February 1999

Seminar für Statistik

Eidgenössische Technische Hochschule (ETH)

CH-8092 Zürich

Switzerland

A sketch of a unifying statistical theory

Frank Hampel

Seminar für Statistik
ETH Zürich
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.

Key words: foundations of statistics, frequentist approach...

1 Overview over the theory

1.1 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., Dempster 1967, 1968, Shafer 1976, Smets 1991), possibility functions (Zadeh 1978, Dubois and Prade 1988), or other "fuzzy" concepts (Zadeh 1965), 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 Smith 1961.) 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 (Hampel 1993a, 1998b).

A closely related concept is that of "cautious surprises" (Hampel 1993a), 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 (1763), who was basically a frequentist, in the light of the new theory (Hampel 1989; cf., e.g., Hampel 1998b).

1.2 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. Hampel 1997a, 1998b). 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" (Hampel 1993a, 1998b), and amazingly enough, this definition works. However, we have to give up full conditionality and hence coherence (Walley 1991); 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. Hampel 1998b). 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 Hampel (1998b).

1.3 Fiducial probabilities

A surprising side result was the clarification of what is correct about the fiducial argument (Fisher 1930, 1956; for a nice historical survey, see Zabell 1992). Fiducial probabilities have been grossly misunderstood by almost everybody, including Fisher himself (for a remarkable exception, see Pitman 1957). For example, in the simplest case, 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 hidden by the notation), is entirely different. Fiducial probabilities are only epistemic, and not the parameter is random, but the whole statement in which it occurs. Fiducial probabilities are derived as a special case of a sideline of the new theory; they are thus put into a much broader framework, which also bridges the gap between continuous and discrete distributions (Hampel 1993a).

1.4 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 Smets 1993), but appears to be true for decisions. Thus, in vague analogy to Smets' "pignistic transformation" (Smets 1990), 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. Hampel 1993b, 1998b). 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.

1.5 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. Pearson 1920, 1921) have already been worked out (Steiner 1995, Hampel 1997b, 1998b). 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, epsilon-successful and epsilon-Bayes solutions might conceivably contain a marked overall improvement over the "pure" classes.

1.6 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" (Hampel 1996; for the background see Hampel et al. 1986, Ch. 1 and Ch. 8.1, and Huber 1981).

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 (Hampel 1993a).

For introductory reading on the theory, the three main papers are Hampel 1993a, 1996 and 1998b; compare also Hampel 1995 and 1997a for a wider horizon. Hampel 1993a is the broadest paper on the theory, but naturally least advanced; Hampel 1996 provides some further solutions which have not yet been followed up; and Hampel 1998b, probably the most readable of the three, pushes farthest what so far has been the main line of development.

2 Remarks on the previous papers about the theory

2.1 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” (Hampel 1989, reproduced in Hampel 1991a; Hampel 1993a), not to be confused with the “possibility” by Zadeh (1978) and Dubois and Prade (1988). It has some esthetic reasons in its favor, but after Hampel (1993a) 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.

2.2 Successful bets and related concepts

The next concept was the key concept of successful bets (Hampel 1990, cf. also Hampel 1991b, both reproduced in Hampel 1991a). It led to a rich and multifaceted paper (Hampel 1993a) 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 (1953/54) 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. Huber 1973; 1981, p. 263f; and Berger’s (e.g. 1984) 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 (Hampel 1993a, 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 Hampel 1990, 1991a, b) contain much material which apparently has to be repeated again and again.

The “surprise” (Hampel 1993a, p. 131, cf. also Hampel 1991a, b), closely related to entropy, is not the only one in the literature. For example, Good (1971; the version in Good 1983 is incomplete), in the spirit of a pure mathematician, defines a whole class of “surprises” whose (according to Good) most important member differs from the above concept just by an additive constant.

The theorem on the relation between cautious surprises and successful bets (Hampel 1993a, 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 (Hampel 1996), 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 $n = k = 2$ (Hampel 1993a, p. 134), Steiner (1995) 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.

2.3 Enforced fair bets

The next major research step was the definition of "best enforced fair bets," a solution for the decision problem, in Hampel (1993b). 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 Hampel 1998b.) 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 Hampel 1998b) 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 (Smets 1990), 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 (1979) "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" (Hampel 1989), is not optimal in the present context.

2.4 Successful bets in different parametric models

The second somewhat larger paper (Hampel 1996) 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., Hampel et al. 1986) and uses a nonstandard type of asymptotics (cf. Hampel 1998a for some general critical comments on asymptotics).

2.5 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 (Hampel 1998b, cf. also Steiner 1995 and Hampel 1997b). 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. Hampel 1993a 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 (1995) 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 (1996) and especially by Coolen (1998) 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

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances, *Philos. Trans. Roy. Soc. London* **A 53**: 370–418. *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.*

Berger, J. O. (1984). The robust Bayesian viewpoint, *in* J. B. Kadane (ed.), *Robustness of Bayesian Analyses*, Elsevier Science, Amsterdam.

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

Choquet, G. (1953/54). Theory of capacities, *Ann. Inst. Fourier* **5**: 131–295.

Coolen, F. P. A. (1998). Low structure imprecise predictive inference for Bayes' problem, *Statistics and Probability Letters* **36**: 349–357.

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

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

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

Fisher, R. A. (1930). Inverse probability, *Proc. Camb. Phil. Soc.* **26**: 528–535.

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

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

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

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

Hampel, F. (1990). Fair bets, successful bets, and the foundations of statistics, *Abstract*, Second World Congress of the Bernoulli Society, Uppsala.

Hampel, F. (1991a). Some remarks on the foundations of statistics, *Research Report 66*, Seminar für Statistik, ETH Zurich.

Hampel, F. (1991b). Some remarks on the foundations of statistics, *Bull. of the ISI, Proc. 48th Session, Volume LIV, Book 4(B), Cairo*, pp. 553–554.

Hampel, F. (1993a). Some thoughts about the foundations of statistics, *in* S. Morgenthaler, E. Ronchetti and W. A. Stahel (eds), *New Directions in Statistical Data Analysis and Robustness*, Birkhäuser Verlag, Basel, pp. 125–137.

Hampel, F. (1993b). Predictive inference and decisions: Successful bets and enforced fair bets, *Proc. 49th Session of the ISI, Contrib. Papers, Book 1. Firenze (Italy)*, pp. 541–542.

Hampel, F. (1995). Zur Grundlagendiskussion in der Statistik, *in* J. Frohn, U. Gather, W. Stute and H. Thöni (eds), *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, Vandenhoeck & Ruprecht, Göttingen, pp. 131–148.

Hampel, F. (1996). On the philosophical foundations of statistics: Bridges to Huber's work, and recent results, *in* H. Rieder (ed.), *Robust Statistics, Data Analysis, and Computer Intensive Methods; In Honor of Peter Huber's 60th Birthday*, number 109 in *Lecture Notes in Statistics*, Springer-Verlag, New York, pp. 185–196.

Hampel, F. (1997a). 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.

Hampel, F. (1997b). How different are frequentist and Bayes solutions near total ignorance?, *Proc. 51th Session of the ISI, Contrib. Papers, Book 2, Istanbul*, pp. 25–26.

Hampel, F. (1998a). Is statistics too difficult?, *Canad. J. Statist.* **26**(3): 497–513.

Hampel, F. (1998b). On the foundations of statistics: A frequentist approach, *in* M. S. de Miranda and I. Pereira (eds), *Estatística: a diversidade na unidade*, Edições Salamandra, Lda., Lisboa, Portugal, pp. 77–97. *Invited talk at the V Congresso Anual, Sociedade Portuguesa de Estatística, Curia, Portugal. June 11 to 14 1997.*

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

Huber, P. J. (1973). The use of Choquet capacities in statistics, *Bull. Internat. Statist. Inst.,* **45**(4): 181–191, (with discussion).

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Pearson, K. (1920). The fundamental problem of practical statistics, *Biometrika* **13**: 1–16.

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

Pitman, E. J. G. (1957). Statistics and science, *J. Amer. Statist. Assoc.* **52**: 322–330.

Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N. J.

Smets, P. (1990). Constructing the pignistic probability function in a context of uncertainty, *in* M. Henrion, R. D. Shachter, L. N. Kanal and J. F. Lemmer (eds), *Uncertainty in Artificial Intelligence*, Vol. 5, Elsevier Science Publ., pp. 29–39.

Smets, P. (1991). 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 (eds), *Uncertainty in Artificial Intelligence*, Vol. 6, Elsevier Science Publ., pp. 375–383.

Smets, P. (1993). 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.

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

Steiner, R. (1995). *Erfolgreiche Wetten für die Binomialverteilung*, Diplomarbeit, Seminar für Statistik, Swiss Federal Institute of Technology (ETH) Zurich.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London.

Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles, *J. R. Statist. Soc. B* **58**(1): 3–34. *With discussion pp. 34–57.*

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

Zadeh, L. A. (1965). Fuzzy sets, *Inform. Control* **8**: 338–353.

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