BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 83, Number 4, July 1977

BOOK REVIEWS

A mathematical theory of evidence, by Glenn Shafer, Princeton Univ. Press, Princeton, New Jersey, 1976, xiii + 297 pp., $17.50 (cloth) and $8.95 (paper).

This is an aptly titled effort to supplement probability theory as developed for chance/aleatory devices by a parallel, but distinct, epistemically oriented quantitative theory of evidence for, and evidential support of, our opinions, judgements of facts, and beliefs. That probability takes its meaning from and is used to describe such diverse phenomena as propensities for physical behavior, propositional attitudes of belief, logical relations of inductive support, and experimental outcomes under prescribed conditions of unlinked repetitions, has long been the source of much of the controversy and vitality in the development and application of probability theory and its associated concepts. Ian Hacking in his recent book *The emergence of probability* [**1**] attempted to trace and explain this intertwining of belief/knowledge and physical (objective) behavior in terms of a conceptual transformation of the categories of knowledge and opinion that was mainly completed by the early 18th century. Hacking's historical/philosophical analysis aims to explain what he holds to be our present dualistic conception of probability as being jointly epistemic (oriented towards assessment of knowledge/belief) and aleatory (oriented towards the objective description of the outcomes of 'random' experiments) with most of the present-day emphasis on the latter. Historically, however, the epistemic component was initially dominant in conceptions of probability.

Probability through the Renaissance applied only to opinions/beliefs and was based upon authoritative testimony in support of these opinions/beliefs. The 19 year-old Leibniz writing in 1665 wished to formalize the evidential support for beliefs by a numerical assignment on a scale of [0, 1] of what he referred to as 'degrees of proof'. The object of this exercise was to be a rationalized jurisprudence. Key to such assignments was an analysis into equally possible (likely) cases.

The growth of an aleatory notion of probability concerning inductive relations between physical signs and physical phenomena starts in the Renaissance. The extent to which the aleatory notion was dependent upon the epistemic notion (there was also a strong converse dependence) is apparent in the posthumously published (1713) *Ars conjectandi* of J. Bernoulli. In Part IV of the *Ars* [**2**] we find the first statement and proof of a law of large numbers, the first firm step on the road to the frequentist/aleatory concepts dominant today. Significantly though, J. Bernoulli was not a frequentist. For Bernoulli, frequency of occurrence was only a clue to the enumeration of the equally possible cases that was the basis of quantitative epistemic probability. Much of Part IV is given over to discussions of evidence and evidential support and how to deal with pure and with mixed evidence; pure evidence either confirmed or disconfirmed the hypothesis. Bernoulli's analysis of this situation led him to be willing to assign probabilities $P(A)$ to hypothesis A and $P(A^c)$ to false A that violated the usual assumption of $P(A) + P(A^c) = 1$, although Bernoulli, as Leibniz, accepted $0 \leqslant P(A) \leqslant 1$.

While there have been sporadic efforts to deal with the problems of evidence and evidential support since J. Bernoulli, these efforts have generated little momentum. The explanation for the slight impact of these attempts to advance probabilistic reasoning seems to lie in the sociology/psychology of mathematics and philosophy and lacks any substantial intellectual basis. Shafer's book is a welcome contribution to the effort to continue the Leibnizian-Bernoullian line, and redress the intellectual imbalance that has developed over the last 200 years by reintroducing issues of practical and intellectual importance for inductive inference. Shafer's approach to the characterization and quantification of evidential reasoning follows suggestions advanced by Dempster [3] and should appeal to the mathematical community as it is a self-contained mathematical theory of evidence, related to Choquet's study of alternating and monotone capacities, that can be viewed as a generalization of probability theory. A relation of this theory to parameter estimation is sketched in Chapter II.

Several of the terms basic to Shafer's discussion of evidence are the following.

(a) Frame of discernment Θ-counterpart to a sample space. List of possibilities relative to our knowledge with distinctions based on our interests. Not a logically exhaustive list descriptive of our best resolving power concerning possibilities. In regard to expanding Θ, Shafer (p. 276) notes "it is always possible to enlarge a frame so as to reduce one's evidence to a collection of nullities". Θ is taken to be finite throughout the discussions.

(b) Basic probability function $m\colon 2^\Theta \to [0,1]$, subject to $m(\phi) = 0$, $\sum_{A \subset \Theta} m(A) = 1$. $m(A)$ reflects the degree of belief exactly committed to A.

(c) The degree of belief or support function Bel: $2^\Theta \to [0,1]$ and satisfies:

$$\mathrm{Bel}(\Theta) = 1;\ \mathrm{Bel}(\phi) = 0;\ (\forall n)(\forall A_1, \ldots, A_n \subset \Theta)\ \mathrm{Bel}\Big(\bigcup_{i=1}^{n} A_i\Big)$$

$$\geqslant \sum_{k=1}^{n} \Big\{ \sum_{i_1 < i_2 < \cdots < i_k} (-1)^{k+1}\, \mathrm{Bel}\Big(\bigcap_{j=1}^{k} A_{i_j}\Big) \Big\}.$$

Bel is a set function that is monotone of order infinity. Bel relates to m through $\mathrm{Bel}(A) = \sum_{B \subset A} m(B)$.

(d) Bayesian belief function is one for which m is positive only on singleton sets and thus probability measure.

(e) Degree of plausibility or upper probability $P^*(A) = 1 - \mathrm{Bel}(A^c)$.

In the case of infinite Θ, discussed in Shafer [4], the function m is of less importance and the argument is based on a representation theorem for monotone set functions showing that they are a composition of a probability measure and an intersection homomorphism.

This framework enables Shafer to reasonably formalize a total absence of relevant evidence bearing on a frame of discernment Θ through the belief function

$$\mathrm{Bel}(A) = \begin{cases} 0 & \text{if } A \neq \Theta, \\ 1 & \text{if } A = \Theta. \end{cases}$$

This characterization of ignorance is preferable to any that has been attempted in the usual setup of probability theory. In a probability setup the only alternatives seem to be to either take no position (e.g., invoke an unknown, as distinct from random, parameter), or to assign a uniform distribution to the elements of Θ, a device with well-known problems.

Central to the theory Shafer develops is a rule of combination of belief functions that appeared in Dempster and a special case of which is credited to J. H. Lambert (1764). From two belief functions Bel_1, Bel_2 on a frame Θ, with associated basic probability functions m_1, m_2 we can form the combined belief function Bel_{12} on Θ with basic probability function m_{12} through

$$\sum_{\{(A_i,B_j):A=A_i\cap B_j\}} m_1(A_i)m_2(B_j) \Big/ \sum_{\{(A_i,B_j):A_i\cap B_j\neq\phi\}} m_1(A_i)m_2(B_j)$$

$$\mathrm{Bel}_{12}(A) = \sum_{B\subset A} m_{12}(B).$$

This rule of combination is applicable when the component belief functions are (p. 57) "based on entirely distinct bodies of evidence" and "the frame of discernment discerns the relevant interaction of the bodies of evidence". Much of the text concerns the mathematical implications of this definition of combination. In terms of it Shafer defines conditional belief functions $\mathrm{Bel}(A|B)$ and assessments of evidence $w(A)$.

A conditional belief function $\mathrm{Bel}(A|B)$ is viewed as the combination of a belief function $\mathrm{Bel}(A)$ and the degenerate belief function

$$\mathrm{Bel}_B(A) = \begin{cases} 1 & \text{if } A \supset B, \\ 0 & \text{if other.} \end{cases}$$

Equivalently, if $P^*(A|B) = 1 - \mathrm{Bel}(A^c|B)$ then

$$P^*(A|B) = \frac{P^*(A \cap B)}{P^*(B)}.$$

Dempster in [3] presents several different definitions of what amounts to $\mathrm{Bel}(A|B)$, his preferred one being the one Shafer adopts.

The assessment of evidence function $w\colon 2^\Theta \to [0, \infty]$ is meant to measure the weight of evidence pointing to any subset of the frame Θ. An elementary belief function S_B, called a simple support function, is defined by

$$S_B(A) = \begin{cases} 1 & \text{if } A = \Theta, \\ s & \text{if } A \supset B, A \neq \Theta, \\ 0 & \text{if } A \text{ other,} \end{cases}$$

where set B is called its focus and $0 < s \leqslant 1$. The corresponding weight of evidence function w_{S_B} is then argued to be given by:

$$w_{S_B}(A) = \begin{cases} \infty & \text{if } A = \Theta, \\ -\log(1 - s) & \text{if } A \supset B, A \neq \Theta, \\ 0 & \text{if } A \text{ other.} \end{cases}$$

Curiously, the Bayesian belief functions then turn out to be pathological in that they can be viewed as arising in the limiting case of infinite contradictory weights of evidence pointing to the atoms of Θ.

While Shafer provides a number of homely examples illustrative of the definitions and their consequences, and some philosophical/interpretive discussion concerning the nature and typology of evidence collections, none of it elaborates how we are to transfer from an evidence collection to the numerical assessments of support, weight of evidence, or degree of belief. Perhaps the absence of elaboration is purposeful, for as Shafer remarks at the close (p. 285), "The construction of a frame of discernment is a creative act . . . The translation of our vague and amorphous knowledge and experience into degrees of support within our frame of discernment can be challenge to the reason and judgement of our astutest minds." The title of this work accurately reflects its mathematical emphasis and its concern with explicating the formal structure of numerical measures of evidential support, albeit Shafer also believes, and I agree, that numerical measures are an idealization. However, a purely mathematical treatment of this subject may be premature if it precedes a sound intuitive grasp of this complex and significant problem. I do not hold with confirmed personalists who might maintain that the relation between a quantitative measure of belief and the basis for this belief is intuitive, primitive, and *a priori.*

It is particularly important that the notion of distinct or separate bodies of evidence be clarified as it is the basis for the essential operation of combining belief functions. The situation here is analogous to that of stochastic independence in probability theory. Stochastic independence is an essential notion of unlinkedness or the uninformativeness of one outcome about another. While it has been explicated mathematically, via the probability of a joint event formed from independent events being equal to the product of their individual probabilities, the adequacy of this explication of our intuitive concept has been questioned [5], and the importance of this issue has been noted by Kolmogorov [6] when he said ". . . one of the most important problems in the philosophy of the natural sciences is . . . to make precise the premises which would make it possible to regard any given real events as independent." Dempster's approach to the combination of bodies of evidence better illustrates this parallel between stochastic independence and distinct bodies of evidence.

This issue of the nature of separate bodies of evidence and the Dempster combination rule also impacts on Shafer's selection of a definition of conditional degree of belief. There is evidently a philosophical issue here as to whether in conditioning on a proposition B we need to think of the knowledge (possibly hypothetical or even counterfactual) that B is true as being based on a separate body of evidence from that which went into the belief function we are conditioning. At any rate this issue is glossed over in the usual probabilistic approach to conditional probability.

The significance and obscurity of the notion of distinct bodies of evidence is also brought out by the possibility of having two distinct bodies of evidence which individually give rise to the same belief function. Yet when we combine the bodies of evidence, they give rise to a different belief function; e.g. if the original basic probability function m was such that $m(A) > 0$, and $m(B) = 0$ if $B \subset A$, then the new function m' may now be positive on subsets of A. The remarks in §8.2 bear on this issue.

Furthermore the issue skirted by renormalizing the joint basic probability function m_{12} to account for the seeming assignment of support to the impossible proposition (ϕ) suggests a defect in the rule of combination. At first reflection an ideal combination rule would not attempt to provide support for ϕ and then have to be adjusted to eliminate this possibility. Admittedly, from the perspective of $P^*(A|B)$ this problem seems less important.

The matter of a decision-making role for belief functions is not addressed. Some discussion of inference, wherein likelihoods are converted to belief functions, is provided in Chapter 11. However, this discussion is flawed (e.g. 11.3), suggesting that the author has not pursued the issue of the utilization of belief functions as closely as he has that of the mathematical characterization of belief functions.

Nonetheless, Shafer's *A mathematical theory of evidence* is a lucid introduction to the unfortunately neglected study of epistemic probability and evidential reasoning. While it is clear that the relations between evidence and beliefs and the classification of types of evidence are more complex than yet accounted for by any formal theory, he at least treats these issues more carefully than is done in the standard probabilistic treatment of inference. Other recent attempts to deal with evidence and epistemic probability would include those centered around Carnap's logical probability [7], I. J. Good's many attempts to mathematicize reasoning [8], [9], and Kyburg's epistemological probability [10]. Hopefully, Shafer's worthy effort will stimulate mathematicians and philosophers to expand their efforts until this subject is at least worthy of the attention of lawyers, as Leibniz hoped it would be 300 years ago!

References

1. I. Hacking, *The emergence of probability*, Cambridge Univ. Press, London and New York, 1975.

2. J. Bernoulli, *Ars conjectandi, Pars quarta* (1713); English transl. by Bing Sung, Tech. Report 12, February 1966, Dept. of Statistics, Harvard Univ., Cambridge, Mass.

3. A. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. Math. Statist. **38** (1967), 325–339. MR **34** #6817.

4. G. Shafer, *A theory of statistical evidence*, Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science (W. Harper and C. Hooker, Editors), Reidel, Dordrecht, Holland (to appear).

5. T. Fine, *Theories of probability*, Academic Press, New York, 1973, pp. 90–94.

6. A. Kolmogorov, *Foundations of the theory of probability*, 2nd ed., English transl., Chelsea, New York, 1956, p. 9. MR **18**, 155.

7. R. Carnap, *Logical foundations of probability*, 2nd ed., Univ. of Chicago Press, Chicago, 1962. MR **32** #2310.

8. I. J. Good, *Probability and the weighing of evidence*, Griffin, London, Hafner, New York, 1950. MR **12**, 837.

9. ———, *The probabilistic explication of information, evidence, surprise, causality, explanation, and utility*, Foundations of Statistical Inference (V. Godambe and D. Sprott, Editors), Holt, Rinehart, and Winston, Toronto, Canada, 1971, pp. 108–141. MR **51** #9280.

10. H. Kyburg, Jr., *Logical foundations of statistical inference*, Reidel, Dordrecht, Holland, 1974.

TERRENCE L. FINE

BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 83, Number 4, July 1977

Order and potential resolvent families of kernels, by Aurel Cornea and Gabriela Licea, Lecture Notes in Mathematics, no. 494, Springer-Verlag, Berlin, Heidelberg, New York, 1975, 154 pp., $7.40.

The first title of this book is *Order and potential*. If the nonspecialist reader opens it at any page, just looking for familiar words, he can be sure to see some mention of *order*, and has reasonable chances to find *potentials*, but may wonder whether the use of the latter word has anything to do with newtonian *potential*, harmonic functions and similar things. After all, the word potential has different connotations in different contexts (the military potential of the United States, the industrial potential of Europe) and the recurrent mention of a mysterious "domination principle" might lead to further political misinterpretations. So let me tell first what the subject of the book really is.

We must come back to the early history of the subject. Between 1945 and 1950, H. Cartan proved some fundamental results in classical potential theory, which were rapidly digested, generalized and improved by the French school of potential theory around M. Brelot, G. Choquet and J. Deny. The axiomatic trend had always been felt in potential theory (the use of the old word "principle" to mean "axiom" may be good evidence for it), and anyhow the years 1950 were those of the big axiomatic boom in mathematics. Hence it is entirely natural that the interest shifted from potential *theory* to potential *theories* defined by suitable axioms. Among the interesting features of classical potential theory, the so called *complete maximum principle* came to play a leading role. It can be easily stated and understood, as follows. Let u and v be two newtonian potentials of positive measures λ and μ, and let a be a positive constant. Assume that

(1) $a + u \geqslant v$ on the closed support F on the measure μ corresponding to v.

Then the same inequality takes place everywhere. This is almost obvious. In the open set F^c complement of F, the function $a + u - v$ is super-