Introduction

In the seventeen years between the first (1954) and second (1971) editions of his book *The Foundations of Statistics*, Savage reported a change in the "climate of opinion" about foundations. That change, he said, "would obliterate rather than restore" his earlier thinking about the relationship between the two major schools of statistics that were the subject of his inquiry. What in the early 1950s started out for Savage as a task of building Bayesian expected utility foundations for common, so-called Frequentist statistics – which for Savage included the Fisher-Neyman-Pearson-Wald program that was dominant in the British-American school from the 1930s – revealed itself, instead, to be an impossibility. Contrary principles separated Bayesian decision theory from what practicing statisticians of the day were taught to do. Significance tests, tests of hypotheses, and confidence intervals give quantitative indices, such as confidence levels, that only accidentally and approximately cohere with the Bayesian theory that Savage hoped might elucidate and justify them.

Since the second edition of *The Foundations of Statistics* and Savage's premature death (both in 1971), we have come to understand much better the extent of the conflict between Bayesian and Frequentist statistical principles of evidence. Some limitations in Frequentist methods, which existed only in the lore of practicing statisticians, gained theoretical footing through the Bayesian point of view. Consider the very general problem of how, within the Frequentist program, to deal with conditional inference, e.g., whether or not to condition on an ancillary statistic. From a Bayesian point of view, conditioning on any ancillary does not affect the resulting posterior distribution. Not so within the Frequentist program, where it is (now) widely recognized as an open challenge.

One version of this problem is optional stopping – whether or not to condition on the experimenter's stopping rule and the resulting changes in Frequentist significance levels that makes. This matter was brought to the fore only in the late 1950s and early 1960s, with much attendant surprise. But by now it has lost its air of "paradox" and is just an accepted part of the difference between the two schools. For example, for the second (1986) edition of *Testing Statistical Hypotheses*, Lehmann added a chapter (10) on "Conditional Inference," noting there that the "discussion will be more tentative than in earlier chapters, and will focus on conceptual aspects more than on technical ones" (p. 539). In our concluding essays (3.7 and 3.8) we revisit the topic of optional stopping. We reach a conclusion only slightly modified from the one Savage did: that the line between Bayesian and Frequentist inference regarding optional stopping is sharp, with the sole exception of reasoning from "improper" priors and finitely additive probability.

Why, then, should anyone continue worrying about foundations? Our view is that foundational studies not only provide for clarification across rival methodologies, but also help point the way to improvements. With that as our banner, this volume collects together sixteen of our essays that deal with what we think are six open issues for Bayesian decision theory and statistics, described below. The reader is alerted that we are not interested, primarily, in issues of classification. Trying to draw a fixed line between what is Bayesian and what is not is futile when that boundary is shifting, as we mean for it to be. Instead, we seek to understand better the scope and limitations of current Bayesian theory with the goal of contributing to its positive growth.

Much of our work on decision theory builds upon one or another of the several foundations laid by Savage (1954), deFinetti (1974), and von Neumann and Morgenstern (1947), together with Anscombe and Aumann (1963). All of these authors arrived at a normative theory of decision making, but by somewhat different means. Fishburn (1970) gives an excellent overview of these and other attempts to lay foundations for decision theory. Here we describe these four theories briefly in order to relate them to the work that we present in our essays.

deFinetti assumed that a statement such as "The probability of event E is 0.6" could be interpreted as a willingness to offer (accept) 0.6 units of some currency for the chance to win (lose) 1.0 units if the event E occurs. That is, the agent who makes such a probability statement believes that the gamble is fair that trades away a sure 0.6 units for the chance to win back 1.0 units (for a net gain of 0.4 units) in case E occurs. If the agent wishes to avoid having a finite number of such fair bets add up to a sure loss, then he/she will have to ensure that the probabilities agree with a finitely additive probability measure. However, deFinetti's theory assumes, in effect, that the units for betting are additive: he develops subjective probability from a theory of additive utility. deFinetti's approach is the subject of our essays 2.1, 2.2, and 2.5.

In their famous theory of cardinal utility, von Neumann and Morgenstern used a decision framework in which an agent holds preferences between *lotteries*: where a lottery is specified by a probability distribution over a set of prizes. When preferences satisfy the axioms of this theory (as summarized in essay 1.2), there exists a unique (cardinal) utility according to which the preferences accord with a ranking of lotteries by their expected utility. In this sense, their theory of utility presupposes probability theory to define the very objects of preference. About 15 years later (and coming after Savage's book), Anscombe and Aumann generalized this framework to one where the agent has preferences over *uncertain* options (called *horse lotteries*), options that consist of mappings from a set of states of nature to a set of von Neumann-Morgenstern lotteries. If an agent's preferences among horse lotteries satisfy several axioms (including those from the von Neumann-Morgenstern theory), then there exists a unique pair consisting of a subjective probability over the states of nature and a (cardinal) utility function over the prizes such that the agent's preferences over horse lotteries agree with a ranking by subjective expected utility. The essays in Part 1 together with essay 2.3 stem from our studies of this approach to decision theory.

Savage (1954) started with a set of options that were merely mappings from states of nature to consequences, rather than mappings to von Neumann-Morgenstern lotteries. He did so to avoid assuming that there existed any agreed-upon probabilities, such as those used (in horse lotteries) to define the von Neumann-Morgenstern lotteries. Thus, Savage's theory takes preference as its primary primitive concept and proceeds without supposing existence of an extraneous concept of probability. Thus, his axioms are somewhat different in kind from those of von Neumann and Morgenstern, and Anscombe and Aumann. But he too concludes that preferences among options should coincide with a ranking by expected utility. Essay 2.4 is particularly concerned with Savage's theory of decision making.

Anscombe and Aumann, Savage, and before them von Neumann and Morgenstern, Ramsey, and deFinetti used normative decision

Cambridge University Press
0521649757 - Rethinking the Foundations of Statistics
Joseph B. Kadane, Mark J. Schervish and Teddy Seidenfeld
Excerpt
More information

theory as the bedrock for building their theories of rationality. But theirs was decision theory for the single decision maker. An overarching topic in our work is understanding how the norms for decision making should apply in settings with more than one (rational) decision maker and then tracing out some of the consequences of this turn for Bayesian statistics. In connection with multiple agent decision making, we write about these four themes:

1. cooperative, non-sequential decisions;
2. the representation and measurement of "partially ordered" preferences – a relaxed version of traditional, Bayesian expected utility that we use with more than one decision maker;
3. non-cooperative, sequential decisions, i.e., game theory;

and

4. pooling rules and Bayesian dynamics for sets of probabilities.

In our work on representation and measurement of Bayesian preference specifically, we explore in detail the following two themes:

5. the significance of state-dependent utility for, e.g., Savage's foundational program of separating belief from value based on an agent's preferences for acts;

and

6. the significance of additivity assumptions for probability, relating also to dominance principles for preference, in connection with statistical decisions involving infinitely many states.

Next, we explain how these six themes relate to the selection of essays in this volume.

THEMES 1 AND 2. Can the norms of expected utility theory be extended from the standard domain of a coherent individual decision maker to apply also to a cooperative, decision-making group of coherent agents, subject to the following (weak) Pareto rule?

Weak Pareto Rule. The group collectively strictly prefers option h_1 to h_2 provided that each individual does.

The overly simple answer to the question is no. The answer we give is better summarized in these words: "No, but the Bayesian theory can be relaxed by weakening the assumption that all acts are comparable in an agent's preference, the so-called Ordering assumption. This weakening of the canonical Bayesian position allows for Pareto-efficient, cooperative decision making, and it builds on the traditional Bayesian justifications of expected utility."

In the opening essay (1.1), "On the Shared Preferences of Two Bayesian Decision Makers," we present our argument for why the answer to the question is no. We show that Pareto-efficient, cooperative decision making cannot be coherent by the usual (Anscombe-Aumann) standards of Bayesian, subjective expected utility.

In essay 1.2, "Decisions Without Ordering." we summarize why attempts to relax expected utility by removing the so-called Independence assumption for preference (in effect, removing Savage's postulate P2) will not work – that leads to unacceptable decisions in sequential settings. In this same essay we show how several well-known justifications for expected utility theory, e.g., the Dutch Book argument, can be applied to a theory which, instead, relaxes "ordering" (Savage's postulate P1). The upshot is a representation of belief (given utility) by a convex set of probabilities, and a representation of value (given belief) by a convex set of utilities. The essay concludes with the observation that simultaneous representation of belief and value, in the absence of "ordering," will not lead to a convex set of probability-utility pairs.

For a proper comparison with, for example, Savage's foundational work on Bayesian statistics, the alternative theory requires a representation of preference in terms of subjective probabilities and utilities. This we provide in the lengthy third essay (1.3), "A Representation of Partially Ordered Preferences," which concludes Part 1 of the volume. The absence of convexity of the representing set (of probability-utility pairs) relates to the need we find for using mathematical induction, rather than the familiar and more elegant methods that separating hyperplanes offer when convexity obtains. Our approach in this essay raises several serious issues about the project of measuring and separating personal probability and utility based solely on an agent's preferences for acts. This issue we take up in Part 2.

THEMES 2, 5, AND 6. That coherent subjective probability can be reduced to rational preferences over acts has been the thesis of many illustrious authors over the years. Notable among these are deFinetti, Ramsey, and Savage. The common thesis of much of this work is that preferences are rational if and only if the preferred option has higher expected utility. Each author demonstrates this conclusion in a different way, but they all rely on mathematical devices for separating the probability from the utility. Due partly to repeated warnings by Herman Rubin (1987) that "you cannot separate probability from utility," a number of authors have recently begun to examine the various attempts to perform the separation. Two of our essays, "Separating Probability Elicitation from Utilities" (2.1) and "State-dependent Utilities" (2.2), show that the reductions of rational preference to state-independent expected utility performed by Savage, deFinetti, and Anscombe and Aumann all fail to take into account the relationship between the value (utility) for a consequence of an act and the state in which the act awards that consequence.

For example, suppose that an agent is asked to compare two acts h_1 and h_2 whose consequences differ on two events A and A^c. The agent may already own assets whose values also differ on A and A^c. The usual elicitation methodology relates the relative valuations of h_1 and h_2 to the supposed probability of A. Consider a very simple case in which $h_1(s) = r$ for each state s in A and $h_1(s) = 0$ for each state s in A^c. On the other hand let h_2 be a lottery based on some device that the agent agrees has certain probabilities of producing certain random occurrences. In this case let $h_2(s) = r$ with probability p and $h_2(s) = 0$ with probability $1 - p$, for all states s. By varying the value of p, the preference between h_1 and h_2 can be made to change. The value of p that makes the agent indifferent between h_1 and h_2 is the elicited probability of A. However, this value will equal the agent's probability of A if and only if the conditional expected utility of gaining r (in addition to the current fortune) given A is the same as the conditional expected utility of gaining r given A^c. If the agent has a nonlinear utility function and if the agent's current fortune contains assets that take different values depending on whether A or A^c occurs, then those two conditional expected utilities might not be equal. Essay 2.1, "Separating Probability Elicitation from Utilities," considers this example and others like it in more detail.

A somewhat simpler situation occurs when the consequences themselves vary in value from one state to the next. In "State-dependent Utilities" (2.2), we consider examples like the following. Suppose that the different states of nature correspond to different currency exchange rates. It is possible for an agent to express preferences between acts with payoffs in each of several single currencies that (for

Cambridge University Press
0521649757 - Rethinking the Foundations of Statistics
Joseph B. Kadane, Mark J. Schervish and Teddy Seidenfeld
Excerpt
More information

a single currency) satisfy the "state-independence" axiom of Anscombe and Aumann. This axiom allows the simultaneous determination of a unique probability over states and state-independent utility function over consequences in each currency, separately. However, the "unique" probabilities elicited using different currencies will differ from each other. This is caused by the fact that the utility function is not state-independent in all currencies simultaneously due to the clear dependence of relative values of currencies on the state of nature.

In "Shared Preferences and State-dependent Utilities" (2.3), we return to the central question of Part 1 of this volume, namely, when does there exist a "coherent" Pareto compromise between two "coherent" agents? This time, however, we examine whether a Pareto consensus exists using a generalized version of "coherence": a version of Anscombe-Aumann's subjective expected utility theory. The generalization is obtained by dropping their axiom that is intended to secure a state-independent utility for consequences. That is, in essay 2.3 we investigate whether or not relaxing the assumption that coherent preference over horse lotteries admits a state-independent utility affects the negative result (essay 1.1) on the absence of Pareto-efficient compromises.

The other two essays in Part 2 (2.4 and 2.5) focus on the added complications caused by extending decision theory from cases with finitely many states to those with infinitely many states. This extension is intimately tied to the definition of consequences. For example, an agent might crudely partition the sure-event into finitely many states to do a rough first analysis of the decision problem. Each act then assigns a consequence (or a lottery over consequences) in each state. Later, when the agent realizes that a finer partition is needed, the various "consequences" in each state need to be split up over the finer partition.

For instance, if a state s in the crude analysis is partitioned into $\{(s, i): i = 1, \ldots\}$, the consequence $h(s)$ that an act h awards in state s of the crude analysis might actually not be of constant value on $\{(s, 1), (s, 2), \ldots\}$, the elements of the finer partition. This phenomenon is discussed by Savage under the heading of "Small Worlds." The crude analysis is carried out in a *small world*, and the finer analysis in a *grand world*. If an act h_1 appears to be preferred to act h_2 in every state in a small world, will h_1 necessarily be preferred to h_2 in the grand world analysis? The answer is tied to the degree of additivity of probability. Both Savage and deFinetti wish to avoid mandating that probability be countably additive, thereby allowing the possibility of finitely additive probability. Our essay 2.4, "A Conflict Between Finite Additivity and Avoiding Dutch Book," explores the connection between how the dominance of h_1 over h_2 in small worlds carries over to grand worlds and the additivity of probability.

The final essay in Part 2, "Statistical Implications of Finitely Additive Probability" (2.5), discusses where the distinction between finitely and countably additive probability surfaces in statistical inference. In particular, considerations such as dominance, discussed above, or admissibility are undermined by allowing probability to be finitely additive. On the other hand, finitely additive probability is necessary in order to be able to incorporate many classical (non-Bayesian) methods within the Bayesian framework. The decision of whether or not to mandate countable additivity must weigh the consequences of each decision against each other, which comes as no surprise to anyone familiar with decision theory.

THEMES 3, 4, AND 6. These themes are the subject of Part 3. Here we examine *dynamic* aspects of non-cooperative, multiagent decision making when agents hold shared evidence. The section begins with two papers discussing relations between subjective expected utility theory and game theory. Their motivation is this. When starting on his book project, *The Foundations of Statistics*, Savage speculated that expected utility theory might provide a sound decision theoretic foundation for Wald's theory of statistical decisions, which he based on a minimax choice principle. (Wald's theory was the cutting edge in statistics at the time, in the early 1950s.) It proved otherwise. As Savage discovered, by contrast with subjective expected utility, minimax-loss undervalues observations whereas minimax-regret violates the ordering postulate P1.

A related question arises in connection with simultaneous move game theory. According to the standard von Neumann-Morgenstern analysis, a player in a single-play, two-person zero-sum game "should" play a minimax mixed strategy because only such a strategy has the property that if the opponent knew the player was using that strategy, the opponent would not change from her own minimax mixed strategy. "Subjective Probability and the Theory of Games" (3.1), the first essay in this group, challenges this line of reasoning using Bayesian principles. The argument is that the principal uncertainty a Bayesian faces in a game situation is what the opponent will do. With a subjective opinion about this, it is easy to calculate an optimal strategy for the player. Only when the Bayesian player's belief happens to coincide with the opponent's minimax strategy will his own minimax strategy be optimal. In the typical case of mixed minimax strategies, every mixture of the strategies in his minimax strategy is equally utility maximizing, so there is no particular reason to play a minimax strategy.

The theme that Bayesian analysis of sequential decision problems does not mandate equilibrium behavior is taken up in "Equilibrium, Common Knowledge, and Optimal Sequential Decisions" (3.2). In this context we show, contrary to Aumann, that correlated equilibrium is not a requirement of certain kinds of common knowledge; also, contrary to Bicchieri, that backwards induction is a valid technique for finding optimal strategies.

The third essay in this group, "A Fair Minimax Theorem for Two-Person (Zero-Sum) Games Involving Finitely Additive Strategies," considers the classical von Neumann-Morgenstern game setup in which there are infinitely many strategies available to each player. Using Wald's "bigger integer" game as an example, the essay shows that if one player is limited to countably additive mixed strategies while the other is permitted finitely additive strategies, the latter wins. When both can play finitely additive strategies, the winner depends on the order in which the integrals are taken. Solutions to this conundrum are discussed.

Essay 3.4, "Randomization in a Bayesian Perspective," addresses a persistent puzzle for Bayesians – the proper role of randomization in the theory. The use of randomization as a justification for inference, as was suggested by Fisher, violates (ironically) the very likelihood principle that Fisher introduced and hence is hopeless from a Bayesian viewpoint. But randomization as a design is often used and useful, and yet is hard to understand theoretically. The analysis in this essay points to multiple decision makers as the key ingredient that leads to the attractiveness of randomization. A subsequent paper (Berry and Kadane, 1997) shows an explicit model with multiple decision makers in which it is optimal for a Bayesian experimental designer to randomize.

Another approach to multiple decision makers is to examine when their opinions can be pooled into a single group opinion. One desirable property a pooling method might have is that it be "externally Bayesian," i.e., the same updated pooled group opinion is obtained whether the pooling takes place before or after updating on the basis of a common likelihood function. Essay 3.5, "Characterization of Externally Bayesian Pooling Operators," gives a necessary and sufficient condition for a pooling operator to be externally Bayesian.

Essay 3.6, "An Approach to Consensus and Certainty with Increasing Evidence," examines the long-run dynamics of iterative Bayesian updating for sets of probabilities described in terms of their extreme members. The primary purpose of this work is to examine what happens to a result of Blackwell and Dubins (1962) about the merging of two Bayesian posterior probabilities (under increasing shared evidence) when the "community" of opinions is characterized by different kinds of extreme views. The Blackwell-Dubins paper is, itself, an important generalization of Savage's result about merging with simple *i.i.d.* data. They avoid all but one of Savage's assumptions: that the rival opinions agree on null events.

As a general methodological point, our essay 3.6 notes that there are two distinct roles that topology plays in such results on asymptotic merging: (1) topology is used to determine merging of posterior opinions in terms of the conditions for convergence of distributions, and (2) topology is used to determine the size of the "community" of rival opinions in terms of the conditions for closure of the set of distributions. We show that it is important to mix and match these distinct roles with different topologies, e.g., using the weak-star topology for both does not support merging in the Blackwell-Dubins setting. Contrast this with the important negative finding of Diaconis and Freedman (1986), who, in a (non-parametric) *i.i.d.* setting, use the weak-star topology for both purposes, versus the positive results in the same setting by Barron, Schervish, and Wasserman (1999), who split the jobs between two topologies.

The role of improper prior distributions, or finitely additive prior distributions, is controversial in Bayesian statistics. Some statisticians use improper distributions, especially uniform distributions, as a representation for ignorance (see Kass and Wasserman, 1996, for a review). Others regard this as wasting the opportunity provided by prior distributions to model the opinion of the client. In the context of this subjective view, a restriction to countably additive distributions can be regarded as unnecessarily restrictive (deFinetti, 1974; Kadane and O'Hagan, 1995).

An important consequence of the restriction to countably additive