The Aggregation of Imprecise Probabilities

Robert F. Nau
Fuqua School of Business
Duke University
www.duke.edu/~rnau

Abstract

Two methods are presented for the aggregation of imprecise probabilities elicited from a group of experts in terms of betting rates. In the first method, the experts bet with a common opponent subject to limits on their personal betting stakes, and their individual and aggregate beliefs are represented by confidence-weighted lower and upper probabilities. In the second method, the experts bet directly with each other as a means of reconciling incoherence, and their beliefs are represented by lower and upper risk neutral probabilities—i.e., products of probabilities and relative marginal utilities for money.

Keywords: lower and upper probabilities, confidence-weighted probabilities, expert resolution, consensus, coherence, arbitrage

1 Introduction

The question of how the subjective probabilities of different individuals ought to be aggregated—if at all—has long been controversial. A well-known impossibility theorem (Genest and Zidek 1986) [9] says that there is no satisfactory way to aggregate ordinary, precise probabilities: the aggregated probabilities must either fail to respect unanimity of beliefs or they must fail to follow the Bayesian recipe for prior-to-posterior updating. Indeed, standard Bayesian theory provides little support for the idea of aggregation: according to the usual definition, all subjective probabilities are purely *personal* probabilities. (See Winkler et al. 1986 [26] for a discussion.) It has often been suggested that imprecise probabilities might be more suitable for aggregation than precise probabilities, because they allow room for disagreement among individuals and they provide additional parameters for reflecting the degree of consensus or dissensus. However, aggregation methods for imprecise probabilities have met with theoretical and practical difficulties of their own.

This paper describes two approaches to the aggregation of imprecise probabilities, both of which have firm "quasi-Bayesian" theoretical foundations and circumvent the impossibility theorem for the aggregation of precise probabilities. They also both have implicit game-theoretic elements, which is not coincidental. We suggest that all subjective probabilities are intrinsically intersubjective in nature—that is, they do not represent beliefs that exist *in vacuo*, but rather beliefs that exist in a strategic environment inhabited by other individuals—and it is only through modeling of the intersubjective dimension that aggregation of probabilities can be justified and carried out in a non-arbitrary fashion.

In the first method, *confidence-weighted lower and upper probabilities* are used as the fundamental representation of uncertainty. The confidence weights permit tradeoffs to be made in a systematic way between the imprecise probability judgments of different individuals, and the confidence-weighted representation has an axiomatic foundation that supports the aggregation of opinions through a slight weakening of the property of transitivity (Nau 1992) [15]. In the second method, the parameters of interest are not "true" subjective probabilities but rather *risk-neutral probabilities*—i.e., products of probabilities and relative marginal utilities for money. An individual's risk-neutral probabilities are the apparent probabilities revealed by his acceptance of money bets under conditions of risk aversion (nonlinear utility for money and significant prior stakes in the outcomes of events). The natural way to aggregate risk-neutral probabilities is to simply let a group of individuals bet with each other and observe the market prices that are obtained after reciprocal learning has occurred, risks have been hedged, and arbitrage opportunities have been exploited. (Markets for betting on securities prices, sporting events, and political races are conducted in exactly this fashion.) Both of these approaches permit the construction of a "representative individual," and the behavior of the representative individual is qualitatively identical to the behavior of a single individual. Hence, the representative individual is necessarily quasi-Bayesian. Furthermore, in both approaches, the imprecision in the revealed probabilities is partly due to the incompleteness of beliefs and partly due to irreducible strategic considerations.

2 Aggregation of confidence-weighted probabilities

In one of the most widely-used models of imprecise probabilities (e.g., Smith 1961, Walley 1991) [23] [24], a *convex set of probability measures* is used to represent the beliefs of an individual, yielding lower and upper probabilities for events as well as lower and upper expectations for random variables. This representation of uncertainty is a natural generalization of precise subjective probabilities that is obtained by relaxing the axiom of completeness. But it is not obvious how aggregation should be carried out under this representation. For example, an aggregate measure of uncertainty could be defined either by the union or the intersection of the sets of probability measures of different individuals, but both of these alternatives appear problematic. The union of convex sets of probability measures generally is nonconvex—although of course it can be convexified—and it yields too loose a representation of aggregate uncertainty. As more opinions are pooled, the union can only get larger, and it reflects only the least informative opinions, whereas intuitively there ought to be (at least the possibility of) an increase in precision as the pool gets larger. On the other hand, the intersection of convex sets of measures may be empty if the experts are mutually incoherent, and it generally yields too tight a representation of aggregate uncertainty. As more opinions are pooled, the intersection can only shrink, and it reflects only the most extreme among those opinions, whereas intuitively there should be some convergence to an "average" opinion when the pool gets sufficiently large. Moreover, neither the union nor the intersection provides an opportunity for the differential weighting of opinions, which would be desirable in cases where one individual is considered (either by herself or by an external evaluator) to be better or worse informed than another individual about a particular event under consideration. Some sort of weighted averaging over lower and upper probabilities could be performed in principle, but the theoretical rationale for doing so is unclear.

It has often been pointed out (e.g., by Good 1962) [10] that a question of infinite regress seemingly arises once it is admitted that probabilities are imprecise: if imprecision in probability judgements is measured by lower and upper bounds, why aren't the bounds themselves somewhat imprecise? But this naturally raises further questions about how the higher-order imprecision could be measured in a meaningful way and about whether there would be practical or conceptual benefits in doing so.

A generalization of lower and upper probabilities that yields a meaningful measure of second-order imprecision was given by Nau (1989, 1992) [13] [14]. There, so-called "confidence weights" are attached to lower and upper probabilities, and an individual is permitted to assess more than one lower or upper probability for a given event, at different levels of confidence. This representation of subjective uncertainty follows in a natural way from a further relaxation of the standard axioms of coherent preferences among gambles, in which transitivity is weakened at the same time as completeness is abandoned. It can also be derived as a generalization of de Finetti's operational method of measuring subjective probabilities.

De Finetti (1937, 1974) [1] [2] proposed that an individual's subjective conditional probability ("prevision") for a event E given another event F should be defined as the (presumably unique) number p for which the individual is indifferent to betting on or against E at rate p, conditional on F. That is, p is the number for which he is willing to accept a gamble with payoff $\alpha(E\text{-}p)F$, where α is a small but otherwise arbitrary number, *positive or negative*, chosen at the discretion of an opponent. (Here E and F are used as indicator variables as well as names of events—i.e., the variable E takes the values 1 or 0 when the event E occurs or doesn't occur, respectively.) De Finetti's definition of subjective probability readily admits a generalization in terms of lower and upper conditional probabilities: for given events E and F, let the individual give lower and upper betting rates p and q, respectively, for $E|F$, meaning that he is willing to accept a gamble $\alpha(E\text{-}p)F + \beta(q\text{-}E)F$, where α and β are small *non-negative* numbers chosen at the discretion of an opponent.

Now consider a further extension of de Finetti's elicitation method in which the individual is assumed to have a *finite total betting stake*—i.e., a maximum amount he can afford to lose. Without loss of generality, let the size of the betting stake be normalized to 1. Then the individual is able to qualify each of his lower or upper betting rates by specifying the maximum fraction of his total stake that he is willing to risk in betting at that rate, and that fraction is defined to be the *confidence weight* associated with the betting rate. For example, he might attach a confidence weight of c to a lower probability p for event E, meaning that he will accept a gamble with payoff $\alpha(c/p)(E\text{-}p)F$, subject to the constraint that $0 \leq \alpha \leq 1$. (Note that because of the scaling factor c/p, the maximum loss is equal to αc, which occurs when E=0 and F=1. The constraint $\alpha \leq 1$ then means the maximum loss must be less than or equal to c.) Similarly, if he attaches a confidence weight of c to an upper probability q, this means he will accept a gamble with payoff $\beta(c/(1\text{-}q))(q\text{-}E)F$. More generally, if he simultaneously assigns lower and upper probabilities p_j and q_j with confidence weight c_j to event $E_j|F_j$, for j=1, …, J, this means he will accept a gamble with total payoff:

$$\Sigma_j\, [\alpha_j(c_j/p_j)(E_j - p_j) + \beta_j(c_j/(1-q_j))(q_j - E_j)]F_j,$$

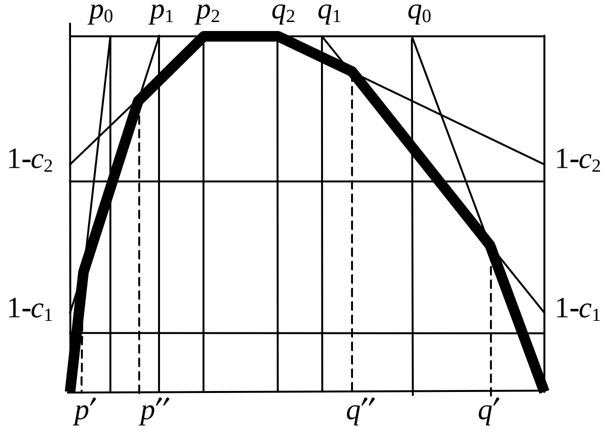
where the nonnegative multipliers $\{\alpha_j\}$and $\{\beta_j\}$are chosen by an opponent subject to the constraint:

$$\Sigma_j\, [\alpha_j + \beta_j] \leq 1$$

(The events $\{E_j\}$ and $\{F_j\}$ in the preceding expression can have any logical relationships whatever, and need not be distinct.) The individual is free to assign more than one confidence-weighted lower or upper probability to the same event—indeed, for reasons to be explained below, he will usually wish to do so. Higher confidence is associated with looser probability bounds: the lesser [greater] of two lower [upper] probabilities will have the higher confidence.

It is intuitively reasonable that, the more favorable the betting rate to himself, the higher the stakes an individual should be willing to play for, but it may not be immediately apparent why he should do so in exactly the way just described—i.e., why he should want to adopt this particular formula for constrained betting at the discretion of an opponent. The rationale for this particular method of attaching a confidence weight to a lower or upper probability is that it directly addresses a shortcoming that de Finetti acknowledged in his original definition of probability, namely "the possibility that people accepting bets against our individual have better information than he has… [which] would bring us to game-theoretic situations" [footnote to Kyburg and Smokler translation]. The constrained betting system described above has the useful property that *it allows the individual to perform a reciprocal measurement on the subjective probability of his opponent,* and to adjust his own betting rate accordingly in advance. To see this, note that the individual's assessment of confidence weighted probabilities presents a well-defined statistical decision problem to a Bayesian opponent ("she"). For simplicity, consider the case of an unconditional probability assessment for a single event E, and suppose that three lower probabilities (p_0, p_1, p_2) and three upper probabilities (q_0, q_1, q_2) are assessed with distinct confidence weights (c_0, c_1, c_2), where $1 = c_0 > c_1 > c_2$. Correspondingly, $p_0 < p_1 < p_2$ and $q_0 > q_1 > q_2$. In other words, the individual is "100% confident" that the probability lies in the interval $[p_0, q_0]$. He is somewhat less confident that it lies in the narrower interval $[p_1, q_1]$, and still less confident that it lies in the still-narrower interval $[p_2, q_2]$. The statistical decision problem that this probability assessment presents to an opponent is described by a *Bayes risk function* (DeGroot 1970) [3] which specifies the opponent's minimum expected loss as a function of her own hypothetical probability distribution. The construction of the Bayes risk function is shown in the graph below.

Figure 1: Construction of the Bayes risk function representing as assessment of confidence-weighted probabilities



Here, the horizontal interval [0,1] represents possible values for the opponent's probability of E.

The vertical interval [0,1] measures the opponent's "loss" which is defined as the opponent's maximum possible gain minus her actual gain. Because the opponent's maximum possible gain is 1 (the size of the probability assessor's total stake), her expected loss is 1 minus her expected gain from the bets she places. Hence her loss is equal to 1 if she places no bets at all, and her loss cannot be less than zero.

Each of the sloping construction lines in the figure is an upper bound on the opponent's Bayes risk function determined by a single confidence-weighted lower or upper probability. In particular, the assessment of a lower probability p with confidence c determines a construction line through the points (0, 1-c) and (p,1). If the opponent takes this bet alone (i.e., assigns it a multiplier α=1), then the opponent's expected loss as a function of her probability is a point somewhere on this line. (If her probability is 0, then she is certain E will not occur, and she expects to win the amount c, in which case her "loss" is equal to 1-c. On the other hand, if her probability for E is exactly p, she thinks the bet is fair, and her expected gain from it is zero, in which case her loss is equal to 1. Expected losses for other probability values follow by linear interpolation.) Similarly, the assessment of a single upper probability q with confidence c determines a construction line through the points (1, 1-c) and (q, 1). If the opponent takes *none* of the bets, her actual and expected loss is equal to 1 regardless of her probability for E, which corresponds to the horizontal line at y=1 in the figure. The opponent's optimal decision is to take the *single* bet, if any, that yields the *minimum expected loss*, and consequently her Bayes risk function is the envelope (pointwise minimum) of the aforementioned construction lines, which is the heavy line shown in the figure.

Note that the Bayes risk function partitions the probability interval into subintervals in which the same bet is always taken by the opponent. For example, if the opponent's probability lies in the subinterval [p', p''], she will take the bet determined by lower probability p_1 with confidence c_1. If the opponent's probability lies in the subinterval [p'', p_2], she will take the bet determined by lower probability p_2 with confidence c_2. (Here p'' is the probability value at which the construction lines corresponding to p_1 and p_2 intersect, etc.) If her probability is in the subinterval [p_2, q_2], she will not bet at all. If her probability is in the subinterval [q_2, q''], she will take the bet determined by upper probability q_2 with confidence c_2, and so on. Now consider how this optimal strategy for the opponent effectively allows the probability assessor to adjust his betting rates in response to what he learns about the opponent's probability. If the opponent's probability is between q_2 and p_2 (the assessor's greatest lower and least upper probability with non-zero confidence), he does not wish to bet at all with the opponent: he cannot discern a profitable difference of opinion. If the opponent's probability is less than p_2 but greater than p'', the assessor is willing to bet at rate p_2. But if the opponent's probability is revealed to be less than p'', though greater than p', the assessor revises his betting rate downward from p_2 to p_1, and so on. Similarly, on the upper end, the assessor is willing to bet against the event at rate q_2 if the opponent's probability is between q_2 and q'', but he revises the rate upward to q_1 if the opponent's probability is revealed to be greater than q'', though less than q'.

In the more general case of multiple, conditional events that are subsets of a finite state space, the Bayes risk function is defined on the probability simplex in $\Re^m$, where m is the number of states. The Bayes risk function on the simplex is typically a piecewise linear concave function—i.e., its graph is a convex polytope. Details are given in Nau (1989, 1992) [14] [15]. The Bayes risk function on the simplex can then be "marginalized" to obtain concave, piecewise linear Bayes risk functions for individual conditional probabilities. These Bayes risk functions behave very much like fuzzy-set membership functions—in particular, they obey the max-min rules of union and intersection—although they are not derived from the assumptions of fuzzy set theory. Rather, these results can be interpreted to establish *as a theorem* that fuzzy set theory is applicable to the modeling of imprecise subjective probabilities, as previously suggested by Freeling (1980) [6] Watson et al. (1979) [25], and Dubois and Prade (1989) [5]. The Bayes risk function also has the same qualititative properties as the "epistemic reliability function" proposed by Gärdenfors and Sahlin (1982, 1983) [7] [8], insofar as it serves to index nested convex sets of probability measures.

The use of confidence-weighted probabilities in decision analysis is discussed by Nau (1989) [13]. Briefly, the second-order imprecision in the probabilities leads naturally to second-order imprecision in expected values of decisions, and the Bayes risk function provides a metric in terms of which alternative decisions can be ranked according to their "distance" from the set of potentially optimal (expected-value-maximizing) decisions. This decision-ranking criterion is similar in principle to methods of sensitivity analysis developed by Rios Insua (1990) [19].

To determine the implications of the confidence-weighted probability model for the aggregation of opinions of different experts, consider a "roomful of experts", each of whom has a finite total betting stake. Let w_i denote the betting stake of expert i, and without loss of generality assume that $\Sigma_i w_i = 1$. Let each expert assess his own subjective uncertainty for various (possibly conditional) events in terms of confidence-weighted probabilities, where the confidence weights are the fractions of his own betting stake that he is willing to risk in betting at the corresponding lower or upper probabilities. (The experts need not all contemplate the same events, although we assume that the events considered by different experts are somehow logically related—i.e., they are subsets of a common state space.) Now consider the statistical decision problem that is perceived by an opponent who bets against all of the experts simultaneously. Because the opponent's gain is naturally the sum of her gains in the bets against the different experts, it immediately follows that her Bayes risk is a weighted average of the Bayes risks of the different experts, with weights equal to their respective betting stakes w_i. In other words, *the Bayes risk function of the combined experts is a linear pool of their individual Bayes risk functions, and the experts are weighted in proportion to the size of their betting stakes.* Because the concavity and piecewise linearity and 0-1 range of the Bayes risk function are preserved under weighted averages, it follows that the Bayes risk function of the combined experts has exactly the same qualititative properties as the Bayes risk function of a single expert. Hence, if all she observes are the combined bets that are available, as summarized by the aggregate Bayes risk function, the opponent cannot really tell whether there is a single expert or multiple experts "on the other side of the door." In particular, if the experts are unanimous in their assessments, then the combined assessment agrees with all their individual assessments as it should. But meanwhile, the combined assessment follows the usual Bayesian rules of prior-to-posterior updating—as generalized to the case of confidence-weighted probabilities—so it is also "externally Bayesian." The impossibility theorem for the aggregation of beliefs thus collapses when beliefs are represented by confidence-weighted probabilities. (See Nau 1990 for further details.) [14]

When confidence-weighted probabilities are aggregated as described above, it is quite possible that the aggregate assessment will be *incoherent*. This will happen, for example, if one expert assesses a lower probability (at some positive level of confidence) which exceeds an upper probability for the same event (at some positive level of confidence) assessed by another expert. However, incoherence is not catastrophic in the framework of confidence-weighted probabilities: the aggregate set of probability measures is not merely empty, as it would be in the case of incoherent precise probabilities or incoherent interval probabilities. Rather, the [fuzzy] set of aggregate probability measures merely has a "subnormal" Bayes risk [membership] function—i.e., a Bayes risk function whose maximum value is less than 1. The common betting opponent will perceive an arbitrage opportunity, but the magnitude of the arbitrage profit—which measures the *relative incoherence* of the assessment—will be bounded at some fraction of the maximum betting stake. In particular, the arbitrage profit is precisely the amount by which the maximum value of the aggregate Bayes risk is less than 1 (Nau 1989) [13]. This measure of relative incoherence is a generalization of one of the two measures of incoherence independently proposed by Schervish et al. (1998) [20]: their model is a special case of confidence-weighted probabilities in which a point probability (i.e., a lower and upper probability that coincide) is assessed for every event with a confidence weight of 1.

The aggregation of confidence-weighted probabilities of two equally-weighted experts for the same unconditional event, in terms of their Bayes risk functions, is illustrated by Figures 2 and 3. Note that the combined Bayes risk function (the heavy line in each figure) is simply an average of the Bayes risk functions of the individual experts, and it typically has a smoother appearance. Figure 2 illustrates the case in which the experts are mutually coherent: the combined Bayes risk is "normal" with a maximum height of 1. Figure 3 illustrates the case in which the experts are mutually incoherent. Here, one expert has asserted that 0.4 is an upper probability while the other has asserted that 0.5 is a lower probability, both with positive confidence. The Bayes risk function is therefore subnormal: its maximum height is 0.95, indicating a relative incoherence of 5%.

Figure 2: aggregation of coherent experts

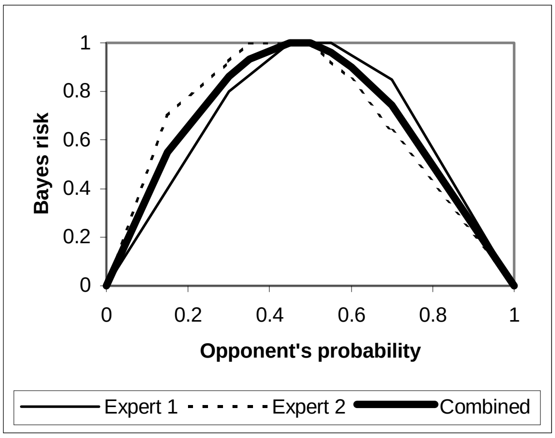
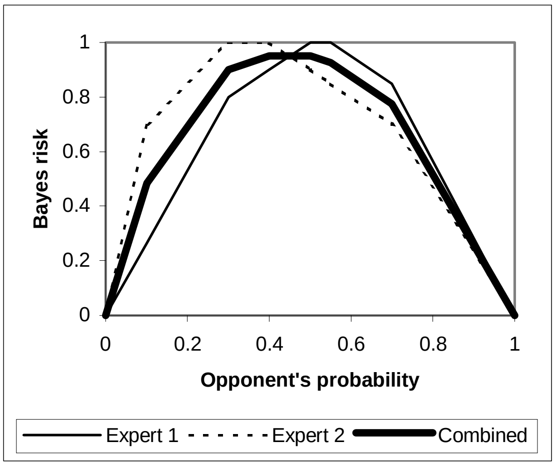


Figure 3: aggregation of incoherent experts



The aggregation method illustrated in the preceding figures can be formalized in terms of linear programming. Suppose that n experts assess confidence-weighted probabilities for the same unconditional event. In particular, suppose that expert i assesses lower and upper probabilities p_{ij} and q_{ij} with confidence c_{ij}, for j=1, …, J_i. Consider the following system of constraints (1), where i ranges from 1 to n and j ranges from 1 to J_i as appropriate:

$$r = \sum_{i=1}^{n} w_i z_i$$

$$z_i \leq 1$$

$$z_i \leq 1 - c_{ij} + (c_{ij}/p_{ij})x \tag{1}$$

$$z_i \leq 1 + c_{ij}q_{ij}/(1\text{-}q_{ij}) - (c_{ij}/(1\text{-}q_{ij}))x$$

$$x \geq 0,\ x \leq 1,\ z_i \geq 0.$$

Next consider the following three linear programs incorporating these constraints:

LP1: maximize r over all $\{r, z_i, x\}$ subject to (1).

LP2: for a fixed x in [0,1], maximize r over all $\{r, z_i\}$ subject to (1).

LP3: for a fixed $\varepsilon \geq 0$, minimize [maximize] x over all $\{r, z_i, x\}$ subject to (1) and $r \geq 1\text{-}\varepsilon$.

Let r^{**} denote the optimal objective value for LP1, let $r^*(x)$ denote the optimal objective value for LP2, and let $x_*(\varepsilon)$ and $x^*(\varepsilon)$ denote the optimal min and max objective values for LP3. Then r^{**} is the maximum value of the Bayes risk function on the interval [0,1], $r^*(x)$ is the value of the Bayes risk function at x, and $x_*(\varepsilon)$ and $x^*(\varepsilon)$ are the endpoints of the level set obtained by cutting the graph of the Bayes risk function at a height of 1-ε. The quantity 1-r^{**} is the relative degree of incoherence of the experts, and $x_*(\varepsilon)$ and $x^*(\varepsilon)$ are the greatest lower and least upper probabilities at a "distance" ε from the "fuzzy" set of aggregate probabilities. (A more efficient and more general LP algorithm applicable to the case of multiple, conditional events is given in Nau 1990 & 1992.) [14] [15]

There are several noteworthy features of this approach to aggregating imprecise probabilities. First, it uses a measure of the *second-order imprecision* in the experts' probability judgments as a basis for making tradeoffs among the lower and upper bounds offered by different experts. The confidence weights provide the necessary measure of second-order imprecision. Second, the assumption of *a limited total betting stake*, and the interpretation of confidence weights as fractions of the total stake allocated to different bets, turns out to be the key to modeling the second-order imprecision and adapting de Finetti's probability elicitation method to the case of multiple experts. Under the usual extension of de Finetti's elicitation method to the case of lower and upper probabilities, there is no explicit limit on the size of any of the bets—they are just said to be "small." If an opponent bets with multiple experts under such conditions, she will only bet against the expert offering the most favorable rate on any given event: she will have no incentive for betting with anyone offering a less attractive rate. Hence the opponent will only "see" the greatest lower and least upper probability bounds among all the experts, and the aggregate set of probability measures will naturally be the intersection of their individual sets of probability measures. Here, by comparison, the fact that each expert's stake is limited means that the opponent will wish to bet some amount against every expert whose "least confident" probability bounds are disjoint from her own probabilities. The possibility that different experts might be endowed with different total betting stakes also provides a basis for differential weighting.

Third, the second-order imprecision has an explicit *intersubjective* dimension: it encodes the way in which the probability assessor wishes to respond to information that is revealed about the probabilities held by his betting opponent. It is precisely this intersubjective quality that sets the stage for a rational aggregation method, because the probability assessments of different experts are all referenced to a common betting opponent.

3 Aggregation of risk-neutral probabilities

In the preceding method of aggregating subjective probabilities, each probability assessor was given the ability to adjust his betting rate in response to the actions of a betting opponent, but the details of the adjustment process had to be specified in advance, encoded in the assignment of appropriate confidence weights to different lower and upper probabilities. The aggregation of such probabilities was imagined to be carried out by assembling a group of experts "in the same room" for purposes of betting against a common opponent, but the experts did not interact directly with each other. The possibility was admitted that the combined experts could be incoherent (i.e., that a lower probability of one expert might be greater than the upper probability of another expert for the same event, at some positive level of confidence), but the effects of incoherence were not disastrous, since betting stakes were *a priori* limited. Indeed, a useful measure of the relative incoherence of the experts was obtained.

An alternative, though conceptually related, approach, is to let the different experts in the same room bet *with each other* prior to interacting with a common betting opponent. Specifically, we assume that upon entering the room, the experts announce to each other their lower and upper betting rates on events—i.e., their bid-ask spreads for Arrow-Debreu securities pegged to those events—and they then proceed to make sequences of small bets with each other at the quoted rates. Presumably their betting rates will change somewhat over time due to a combination of reciprocal learning, mutual hedging of risks, and exploitation of arbitrage opportunities. Under these conditions, an expert's bid-ask spread for a given security at any given time may reflect not only the intrinsic imprecision in his beliefs but also his desire to hedge himself against the possibility that his opponents have superior information and/or to profit from the possibility that his opponents have inferior information. Note that if the individual bets are small relative to the experts' total betting stakes, and if the experts have the opportunity to adjust their rates continuously over time, it is unnecessary for them to use the device of confidence weights described in the preceding section: the only betting rates that matter at any given moment are their greatest lower and least upper betting rates with positive confidence.

In this setting (or indeed, any setting in which material rewards are used to elicit beliefs), the agents' lower and upper betting rates on events are not necessarily measures of pure belief. Rather, they are measures of belief *compounded with possibly-state-dependent marginal utilities for money*. For, on the assumption that an individual is an expected-utility-maximizer, his (greatest lower or least upper) betting rate on an event ought to be proportional to the product of his true (lower or upper) probability and his relative marginal utility for money given the occurrence of that event, notwithstanding any additional hedging that he may wish to perform for strategic reasons. This product of true probability and relative marginal utility for money is known as a *risk neutral probability* in the literature of finance: it is the probability we would infer from the individual's betting behavior on the (possibly incorrect) assumption that he is risk neutral. If the individuals in the room all have decreasing marginal utility for money in every state, then as they accumulate bets with each other, their state-dependent marginal utilities will shift so as to produce a convergence in their risk neutral probabilities, even if their "true" probabilities remain relatively constant. Of course, their true probabilities might also converge to some extent as the experts learn from their interactions with each other, but the precise extent of that convergence will remain unknown: their true probabilities and their marginal utilities will be inseparable in the eyes of an observer. Fortunately, the distortion of subjective probabilities by marginal utilities for money is not catastrophic for purposes of decision analysis. It is possible in principle to perform decision analysis using risk neutral probabilities (Nau 1995a) [16], and risk neutral probabilities also provide the basis for a useful integration of decision analysis and options pricing methods of project valuation (Smith and Nau 1995) [22]. Further discussion of the practical problem of separating probability and utility is given by Kadane and Winkler (1988) [11], Schervish et al. (1990) [20] and Karni and Safra (1995). [12]

As the experts in the room continue to bet with each, updating their beliefs and marginal utilities and strategic positions, an equilibrium will eventually be reached in which their betting rates converge on stable values, and those stable values will necessarily be coherent if all arbitrage opportunities have been rationally exploited. A "consensus" of the experts is thereby achieved, but the consensus is with respect to risk neutral probabilities rather than "true" subjective probabilities of events. In the final equilibrium, every expert has his own lower and upper risk neutral probabilities for events (or more generally, bid-ask spreads for arbitrary securities pegged to those events), defining *a convex set of risk-neutral probability measures*. From the perspective of an observer who enters the room at this point, only the greatest lower and least upper betting rates (or bid-ask prices) are of interest. Hence, from the perspective of the observer, there is a single "representative agent" whose convex set of risk-neutral probability measures is the *intersection* of the (final) sets of risk-neutral probability measures of the separate experts. Thus we find support for the idea of representing the aggregate opinion of the experts by the intersection of their respective convex sets of probability measures, but the probability measures are risk-neutral rather than "true" measures, and the intersection is taken only *after* the experts have had the opportunity to bet with each other and to exploit any arbitrage opportunities that may be discovered. The problem of incoherence (i.e., an empty intersection) therefore does not arise at all, and the problem that only the most "extreme" beliefs are represented is mitigated by the fact that the experts have had the opportunity to learn from each other and revise their beliefs.

As a simple example, suppose there are two experts and a single event E whose probability is of interest. Suppose the experts have initial lower and upper risk neutral probabilities $[p_1, q_1]$ and $[p_2, q_2]$ respectively. If the intervals overlap—say, if $q_1 > p_2$—then the experts will not bet at all with each other, and trivially their aggregated risk neutral probabilities will be the intersection of the intervals, namely $[p_1, q_2]$. In the more interesting case where their intervals are initially disjoint—say, $q_1 < p_2$—they will bet with each other and adjust their risk neutral probabilities until some overlap is achieved. Suppose, for example, that their initial intervals are [0.2, 0.3] and [0.4, 0.5], respectively. Then for any betting rate between 0.3 and 0.4, expert 1 is initially willing to bet against E while expert 2 is willing to bet on E. The actual rate at which betting occurs may depend on which agent has more patience, stubbornness, or power relative to the other. A canonical way to model the outcome of the betting is to imagine that an arbitrageur bets with each expert at the least favorable rate the expert is willing to accept, so that the expert's total expected utility remains unchanged. Thus, initially, the arbitrageur will bet with expert 1 at the rate 0.3 (with the expert betting *against E*) and with expert 2 at the rate 0.4 (with the expert betting *on E*). As the stakes accumulate, the betting rate of expert 1 will rise while that of expert 2 falls, until they converge on a common value that represents their aggregate beliefs. The arbitrageur will reap a positive profit during this process, and the total arbitrage profit can later be redistributed (or not) between the experts in an arbitrary manner, which may (or may not) trigger further rounds of betting. If the arbitrage profits are *not* redistributed, the solution is generally unique (under suitable regularity conditions on the utility functions) and can be found by solving a nonlinear programming problem in which the arbitrageur's minimum profit across states is maximized while holding the experts' expected utilities constant.

To complete the analysis in this example, suppose that the experts have exponential utility functions with risk tolerances of \$10,000 and \$20,000 respectively. A convenient property of exponential utility functions is that they exhibit constant absolute risk aversion, which implies that lump-sum changes in wealth have no effects on risk neutral probabilities. Hence, if the experts have exponential utility functions, they will converge on a unique risk neutral distribution regardless of whether or how they share the arbitrage profits. In this case, with respective risk tolerances of $10,000 and $20,000, they will converge uniquely on an aggregate risk neutral probability of 0.3652, which is closer to the initial risk neutral probability of expert 2 than expert 1 because expert 2 has a higher risk tolerance. (Under more general utility functions, such as log or power functions, the solution would not be unique and would depend on details of the betting sequence and the sharing of profits.) In this example it has been assumed that no learning or strategic maneuvering takes place, so that changes in the experts' risk neutral probabilities are due *only* to changes in relative marginal utilities as betting stakes accumulate. If their true probabilities also change because they learn from each other while betting, the final solution will be hard (perhaps impossible) to predict in advance.

The convergence of risk neutral probabilities among agents who are free to bet with each other, leading to the construction of a representative agent, is discussed by Kadane and Winkler (1988) [11], Nau and McCardle (1991) [18], and Nau (1995b) [17], and it is also well known to be a necessary condition for competitive equilibrium in a contingent claims market (Drèze 1970). [4] The risk-neutral probabilities are otherwise known as "state prices." Of course, this market-based approach to aggregating risk-neutral probabilities is widely employed in practice, most notably in markets for derivative securities (e.g., stock options and futures) and for betting on sporting events (e.g., parimutuel oddsmaking) and political races (e.g., the Iowa political markets). The efficiency and calibration of such markets are well known.

We see once again that, in order to aggregate subjective probabilities in a rational and non-arbitrary fashion, it is necessary to begin with a definition of probability which is intrinsically intersubjective in nature and which allows for varying degrees of precision in the assessed values. In this case, the intersubjectivity resides in the fact that the experts are given the opportunity to bet with each other during the elicitation process, so that their revealed (risk-neutral) probabilities are not merely measures of private belief but rather are measures of public willingness to bet against each other. The final risk-neutral probabilities to which experts converge under these conditions are not predictable *a priori* from their hypothetical true prior probabilities and utilities, even in principle, except under restrictive conditions (exponential utility, no learning, etc.). There may be many possible final allocations of state-contingent wealth that are mutually preferred to the initial allocation, and correspondingly many possible systems of final state prices. The one to which the experts converge may depend on the vagaries of the sequence of trades, on their relative bargaining powers with respect to each other, on the way they learn from their interactions with each other, and on other psychological factors or environmental contingencies. The final equilibrium therefore is not uniquely determined by initial conditions that are subject to independent measurement. Rather, the observation of the final equilibrium is a *fundamental measurement* of aggregate belief.

Acknowledgements

This research was supported by the National Science Foundation under grant 98-09225, by the Fuqua School of Business, and by INSEAD. This paper was written while the author was a visitor at INSEAD.

References

[1] de Finetti, B. (1937) La Prévision: Ses Lois Logiques, Ses Sources Subjectives. *Ann. Inst. Henri Poincaré* **7**, 1-68. Translation reprinted in H.E. Kyburg and H.E. Smokler, eds. (1980) *Studies in Subjective Probability*, 2nd ed., Robert Krieger, New York, 53-118

[2] de Finetti, B. (1974) *Theory of Probability, Vol. 1.* Wiley, New York

[3] DeGroot, M. (1970) *Optimal Statistical Decisions.* McGraw-Hill, New York

[4] Drèze, J. (1970) Market Allocation Under Uncertainty. *European Econ. Rev*. **2**, 133-165

[5] Dubois, D. and H. Prade (1989) Fuzzy Sets, Probability, and Measurement. *European J. Operns. Res.* **40**, 135-154

[6] Freeling, A.N.S. (1980) Fuzzy Sets and Decision Analysis. *IEEE Trans. Sys. Man & Cyb*. **SMC-10**:7, 341-354

[7] Gärdenfors, P., and N. Sahlin (1982) Unreliable Probabilities, Risk-Taking, and Decision Making. *Synthese* **53**, 361-386

[8] Gärdenfors, P. and N. Sahlin (1983) Decision Making With Unreliable Probabilities. *Brit. J. Math. Stat. Psych* **36**, 240-251

[9] Genest, C. and J.V. Zidek (1986) Combining Probability Distributions: A Critique and Annotated Bibliography. *Statistical Science* **1**, 114-148

[10] Good, I.J. (1962) Subjective Probability as the Measure of a Non-Measurable Set. In E. Nagel, P. Suppes, and A. Tarski (eds.) *Logic, Methodology, and Philosophy of Science.* Stanford University Press. Reprinted in Kyburg and Smokler (1980)

[11] Kadane, J.B. and R.L. Winkler (1988) Separating Probability Elicitation from Utilities. *J. Amer. Statist. Assoc.* **83**, 357-363.

[12] Karni, E. and Z. Safra (1995) The Impossibility of Experimental Elicitation of Subjective Probabilities, *Theory and Decision* **38**, 313-320.

[13] Nau, R.F. (1989) Decision Analysis with Indeterminate or Incoherent Probabilities. *Annals of Operations Research* **19**, 375-403

[14] Nau. R.F. (1990) Indeterminate Probabilities and Utilities on Finite Sets. Fuqua School of Business Working Paper

[15] Nau, R.F. (1992) Indeterminate Probabilities on Finite Sets. *The Annals of Statistics* **20**:4, 1737-1767

[16] Nau, R.F. (1995a) Coherent Decision Analysis with Inseparable Probabilities and Utilities. *Journal of Risk and Uncertainty* **10**, 71-91

[17] Nau, R.F. (1995b) The Incoherence of Agreeing to Disagree. *Theory and Decision* **39**, 219-239

[18] Nau, R.F. and K.F. McCardle (1991) Arbitrage, Rationality, and Equilibrium. *Theory and Decision* **31**, 199-240

[19] Rios Insua, D. (1990) *Sensitivity Analysis in Multi-objective Decision Making*. Springer-Verlag, Berlin

[20] Schervish, M.J., T. Seidenfeld, & J.B. Kadane (1990) State-Dependent Utilities. *J. Amer. Statist. Assoc.* **85**, 840-847.

[21] Schervish, M.J., T. Seidenfeld, & J.B. Kadane (1998) Two Measures of Incoherence: How Not to Gamble If You Must. Technical Report #660, Department of Statistics, Carnegie Mellon University, Pittsburgh PA 15213

[22] Smith, J.E. and R.F. Nau (1995) Valuing Risky Projects: Option Pricing Theory and Decision Analysis. *Management Science* **41**:5, 795-816

[23] Smith, C.A.B. (1961) Consistency in Statistical Inference and Decision. *J. Roy. Statist. Soc.B* **23**, 1-25

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

[25] Watson, S.R., Weiss, J.J., and M.L. Donnell (1979) Fuzzy Decision Analysis. *IEEE Trans Sys. Man & Cyb.* **SMC-9**:1 1-9

[26] Winkler, R.L., D.V. Lindley, M.J. Schervish, R.T. Clemen, S. French, and P.A. Morris (1986) Expert Resolution. *Management Science* **32**, 298-328