COHERENT ASSESSMENT OF SUBJECTIVE PROBABILITY

by

Robert F. Nau
Operations Research Center
University of California, Berkeley

MARCH 1981 ORC 81-5

This research has been partially supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-77-3179 and the Office of Naval Research under Contract N00014-75-C-0781 with the University of California.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ORC 81-5	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COHERENT ASSESSMENT OF SUBJECTIVE PROBABILITY		5. TYPE OF REPORT & PERIOD COVERED Research Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Robert F. Nau		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0781
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042 238
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, Virginia 22217		12. REPORT DATE March 1981 13. NUMBER OF PAGES 49
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-77-3179.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Subjective Probability
Coherence
Admissibility
Scoring Rules
Linear Programming
Upper and Lower Probabilities
Game Theory

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ACKNOWLEDGEMENTS

To Professor Richard E. Barlow for his generous encouragement, advice, and support; and to Professor Dennis V. Lindley for his correspondence while this report was in progress, and for awakening me to the Bayesian viewpoint.

ABSTRACT

This report discusses the role of coherence considerations in the definition and measurement of subjective probability. A general version of De Finetti's coherence theorem--that either a set of betting probabilities obeys the laws of probability or else a sure win is possible for the bettor--is proved, using a variant of Farka's Lemma. This theorem provides the basis for several admissibility theorems for scoring-rule probabilities, under a generalization of scoring rules suggested by Lindley. Linear programming methods for identifying and reconciling incoherence are discussed, and a comparison is made with Bayesian reconciliation methods.

Coherent Assessment of Subjective Probability

by Robert F. Nau

1. Introduction

The purpose of this report is to aggregate and generalize some well-known results of de Finetti (1937, 1972, 1974), Smith (1961), and Savage (1971) and some recent results of Lindley (1980) concerning the use of betting systems and scoring rules for eliciting subjective probabilities, and to discuss methods for identifying and reconciling incoherence. The principal analytic tool will be a separating-hyperplane theorem of linear algebra, which, together with its variants and extensions, has previously been applied by numerous authors to discussions of coherence and admissibility in statistical inference and decision. (E.g., Blackwell and Girshick (1954), Smith (1961), Cornfield (1969), Freedman and Purves (1969), Dawid and Stone (1972, 1973), Heath and Sudderth (1972, 1978), Pierce (1973), and Buehler (1976)). The central problem discussed here is the elicitation of subjective conditional probabilities for a set of events which are subsets of a finite sample space, with conditional probabilities directly defined in terms of "called-off bets," rather than as ratios of unconditional probabilities. Coherence (of betting probabilities) and admissibility (of scoring-rule probabilities) are defined in terms of avoiding unnecessary certain loss under all outcomes in the sample space, and are shown to be equivalent criteria for defining and measuring subjective probability. A general version of de Finetti's coherence theorem, stated in terms of lower and upper conditional betting probabilities, is proved using the separating-hyperplane theorem below. The coherence theorem provides a basis for several admissibility theorems for scoring-rule probabilities, under a generalization of scoring rules suggested by Lindley (1980). Throughout, emphasis is placed on the distinction between strict and non-strict forms of coherence and admissibility, illuminating the role of zero probabilities in subjectivistic theory. The construction of a probability measure consistent with a given set of betting probabilities, whose existence is required for coherence, is shown to be a simple linear programming problem, whose dual is the search for a combination of bets providing a "sure win." The geometric interpretation of coherence suggests linear programming methods for improving the precision of probability assessments and reconciling incoherence, using lower and upper probabilities to characterize imprecise initial assessments. These methods are shown to provide a computationally simpler alternative to the Bayesian reconciliation methods of Lindley, Tversky, and Brown (1979).

The various so-called separating-hyperplane theorems can all be derived from a "basic separation theorem" for linear spaces (Dunford and Schwarz (1958), p. 412), which states that any two disjoint convex sets (say, X and Y), one of which has an interior point, can be separated by a non-trivial linear functional--i.e. there exists a linear functional f, not identically zero, and a real number d such that $Re[f(\mathbf{x})] \leqslant d$ for all $\mathbf{x}$ in X and $Re[f(\mathbf{y})] \geqslant d$ for all $\mathbf{y}$ in Y. If the two sets are also closed, then the separation can be made strict (i.e., strict inequality can be obtained in at least one of the above relations). In finite-dimensional Euclidean space the linear functional takes the form $f(\mathbf{x})=\mathbf{z}'\mathbf{x}$ where $\mathbf{z}$ is a fixed vector, with the geometric interpretation that X and Y are separated by a hyperplane whose normal direction is $\mathbf{z}$ and whose distance from the origin is d. The following conventions and notation for vector inequalities will be useful: a vector $\mathbf{x}$ is *nonnegative* ("$\mathbf{x} \geqq 0$") if all of its components are nonnegative; $\mathbf{x}$ is *semi-positive* ("$\mathbf{x} \geqslant 0$") if it is nonnegative and not the zero vector; and $\mathbf{x}$ is *positive* ("$\mathbf{x} > 0$") if all of its components are positive, where 0 denotes the zero vector of appropriate length. Corresponding definitions and notation apply to *non-positive*, *semi-negative*, and *negative* vectors. In these terms, the theorem for later use is:

THEOREM 1. Exactly one of the following two systems has a solution:

(*i*) $\mathbf{A}\mathbf{z} < 0$ [$\mathbf{A}\mathbf{z} \leqslant 0$], $\mathbf{z} \geqslant 0$

(ii) $\mathbf{w}'\mathbf{A} \geqq 0$, $\mathbf{w} \geqslant 0$ $[\mathbf{w} > 0]$, $\sum_j w_j = 1$

where $\mathbf{A}$ is a matrix and $\mathbf{z}$ and $\mathbf{w}$ are vectors of appropriate length.

Proof: For the unbracketed case, obviously both systems cannot simultaneously have solutions. Then either the non-negative orthant contains a point of the closed convex hull of the row vectors of $\mathbf{A}$, in which case (ii) has a solution, or else there exists a hyperplane which strictly separates these two closed convex sets. The normal direction of this hyperplane constitutes a solution to (i). For the bracketed case, again both systems cannot simultaneously have solutions. Then either the non-negative orthant contains a point of the open convex cone of the row vectors of $\mathbf{A}$, in which case a solution to (ii) is obtained by normalization, or else for some j the following system must have no solution: $\mathbf{w}'\mathbf{A} + \mathbf{a}^j \geqq 0$, $\mathbf{w} \geqslant 0$, where $\mathbf{a}^j$ denotes the j^{th} row vector of $\mathbf{A}$. (If this system had a solution for every j, then their sum plus the vector whose components are all 1's would constitute a solution to (ii) following normalization.) For some j, then, the non-negative orthant has no point in common with the closed convex set formed by the direct sum of $\mathbf{a}^j$ and the closed convex cone of all the row vectors of $\mathbf{A}$, so that these two sets are strictly separated by some hyperplane. The normal direction of this hyperplane then constitutes a solution to (i), in which the j^{th} element of $\mathbf{Az}$ is negative.

In applications, a vector $\mathbf{w}$ satisfying (ii) will be considered to represent a probability distribution. The following corollary is closely related to Farka's Lemma ("either $\mathbf{Az} \leqq 0$, $\mathbf{c}'\mathbf{z}>0$ has a solution, or else $\mathbf{w}'\mathbf{A}=\mathbf{c}$, $\mathbf{w} \geqq 0$ has a solution"), which is the basis of the duality theorem of linear programming.

COROLLARY: Exactly one of the following two systems has a solution:

(i) $\mathbf{Az} < 0$ $[\mathbf{Az} \leqslant 0]$

(ii) $\mathbf{w}'\mathbf{A}=0$, $\mathbf{w} \geqslant 0$ $[\mathbf{w} > 0]$, $\sum_j w_j = 1$

This follows by applying Theorem 1 to the matrix $[\mathbf{A}|-\mathbf{A}]$.

2. Coherence for the fair bookie

This section deals with the elicitation of subjective probabilities under a betting system--that is, a framework in which a transaction takes place between a bookie and a bettor. De Finetti's (1937) well-known theorem on coherence--that either a bookie's betting probabilities ("bet prices") obey the laws of probability or else a sure win is possible for the bettor--is proved for the general case of conditional bets on a finite number of events, and the geometrical interpretation of coherence is discussed. A further generalization of the coherence theorem to incorporate lower and upper probabilities is given in a later section.

Consider a bookie, a bettor, and n pairs of events: (E_i, F_i), $i=1,...,n$. The bookie must establish prices ("set the odds") for bets on E_i conditional on F_i ("E_i given F_i") for all i, and the bettor may then place any combination of bets. Following de Finetti's convention, capital letters such as E and F will be used interchangeably as names for events and also as the indicator variables for the same events-- e.g., "$E=1$" is interchangeable with "E is true", and "$1-E$" is interchangeable with $\bar{E}$ ("not-E"). Let the transaction be described as follows: first the bookie chooses a vector $\mathbf{p}=(p_1, \ldots, p_n)$, where p_i is his price for buying or selling a "unit bet" on E_i given F_i-- i.e., a lottery which pays 1 unit if $E_iF_i=1$, pays zero if $(1-E_i)F_i=1$, and pays back the purchase price (in which case the bet is considered "called off") if $F_i=0$. (The bookie is "fair" in the sense that he buys and sells at the same price. The more general case of unequal buying and selling prices is discussed in a later section.) The bettor then chooses a vector $\mathbf{z}=(z_1, \ldots, z_n)$ where $|z_i|$ is the number of unit bets on E_i given F_i that he wishes to buy (if $z_i>0$) or sell (if $z_i<0$). The net gain to the bookie for the bet on the i^{th} event pair in all cases is given by the expression $(p_i-E_i)F_iz_i$, which may be positive, negative, or zero. In conventional betting parlance, the bookie is said to have offered "odds of $(1-p_i)$ to p_i against E_i" and reciprocal odds "on" E_i; the bettor has placed a stake of $|p_iz_i|$ "on" E_i if $z_i>0$ or "against" E_i if $z_i<0$--all conditional on F_i.

It is assumed that the E's and F's are subsets of a sample space, Θ, consisting of m mutually exclusive and collectively exhaustive outcomes which are known to both the bookie and the bettor. That is, both participants are aware of all logical dependencies among the $2n$ events of interest, which place restrictions on their possible joint realizations, since every possible joint realization must correspond to at least one outcome in the sample space. Let θ_j denote the j^{th} element of Θ, and also the event consisting of only that outcome. Let E_{ij} and F_{ij} denote the values of E_i and F_i under outcome j-- i.e., $E_{ij}=1$ if E_i contains θ_j, and $E_{ij}=0$ otherwise. Then the total net gain to the bookie for all n bets when θ_j obtains is

$$t_j(\mathbf{z};\mathbf{p}) = \sum_{i=1}^{n}(p_i - E_{ij})F_{ij}z_i \ . \tag{2.1}$$

The "payoff vector" for the bookie, $\mathbf{t}(\mathbf{z};\mathbf{p})$, is now defined as the m-vector whose j^{th} element is $t_j(\mathbf{z},\mathbf{p})$.

DEFINITION: The vector of prices $\mathbf{p}$ is [*strictly*] *coherent* for the bookie if-and-only-if there does not exist any vector of bets $\mathbf{z}$ for which the resulting payoff vector is [semi-] negative.

In other words, the bookie's prices are coherent if there is no "sure-win" bet for the bettor (one for which the bookie loses money under every outcome), and strictly coherent if there is no "can't-lose" bet (one for which the bookie loses under at least one outcome, and wins under none). Necessary conditions for coherence or strict coherence in certain cases can be immediately identified. For example, if $F_i=0$ is impossible, then coherence requires $0 \leqslant p_i \leqslant 1$, since choosing $z_i>0$ if $p_i<0$, or $z_i<0$ if $p_i>1$, would produce a sure win for the bettor. Similarly, if both $E_iF_i=1$ and $(1-E_i)F_i=1$ are possible, then *strict* coherence requires $0<p_i<1$. On the other hand, if $E_iF_i=1$ is possible but $(1-E_i)F_i=1$ is not (or vice versa), then strict coherence requires $p_i=1$ (or $p_i=0$). In general, the necessary and sufficient conditions for coherence or strict coherence are given by:

THEOREM 2. $\mathbf{p}$ is [strictly] coherent if-and-only-if there exists a [positive] probability distribution $\mathbf{w}$ on Θ, and a corresponding probability measure $\mathbf{P_w}$ on all subsets of Θ, such that for every i either $p_i=\mathbf{P_w}(E_i|F_i)$ or else $\mathbf{P_w}(F_i)=0$.

Proof: Let $\mathbf{A}$ be the $m \times n$ matrix whose $(j,i)^{\text{th}}$ element is $(p_i - E_{ij})F_{ij}$. Then $\mathbf{t}(\mathbf{z};\mathbf{p}) = \mathbf{A}\mathbf{z}$. By the Corollary to Theorem 1, $\mathbf{A}\mathbf{z}<0$ $[\mathbf{A}\mathbf{z} \leqslant 0]$ has no solution if-and-only-if there exists $\mathbf{w} \geqslant 0$ $[\mathbf{w} > 0]$ such that $\mathbf{w}'\mathbf{A}=0$. This vector equality is equivalent to:

$$\sum_{j=1}^{m} (p_i - E_{ij}) F_{ij} w_j = 0, \; i=1,\ldots,n \tag{2.2}$$

whence either

$$\sum_{j=1}^{m} F_{ij} w_j = 0, \tag{2.3}$$

or else

$$p_i = \frac{\sum_{j=1}^{m} E_{ij} F_{ij} w_j}{\sum_{j=1}^{m} F_{ij} w_j}. \tag{2.4}$$

Let $\mathbf{P}_{\mathbf{w}}$ be the unique, finitely additive probability measure on all subsets of Θ which satisfies $\mathbf{P}_{\mathbf{w}}(\theta_j)=w_j$ for all j. That is,

$$\mathbf{P}_{\mathbf{w}}(F_i) \equiv \sum_{j=1}^{m} F_{ij} w_j \tag{2.5}$$

and similarly for all other subsets of the sample space. Define the conditional probability of E_i given F_i in the usual way as

$$\mathbf{P}_{\mathbf{w}}(E_i | F_i) \equiv \frac{\mathbf{P}_{\mathbf{w}}(E_i F_i)}{\mathbf{P}_{\mathbf{w}}(F_i)}. \tag{2.6}$$

Substitution of (2.5) and (2.6) into (2.3) and (2.4) completes the proof.

This theorem provides the motivation for de Finetti's definition of subjective probabilities as coherent bet prices. From the definition of the probability measure $\mathbf{P}_{\mathbf{w}}$ in (2.5) and (2.6), it follows that the quantities $\mathbf{P}_{\mathbf{w}}(E_i|F_i), i=1, \ldots, n$, obey the usual "laws" of probability, including the additive and multiplicative laws. (This will be illustrated in the geometrical examples below.) Theorem 2 implies that, by conformity with some such measure, coherent bet prices obey the same laws merely through the fact of being coherent, rather than by prior assumption. Thus, if coherence is taken as an axiom of subjective probability, the probability laws which are traditionally stated as axioms or definitions are obtained instead as theorems. (De Finetti

(1937) gives separate proofs of the "total probability theorem," or additive law, and the "compound probability theorem," or multiplicative law, based on evalution of the determinant of **A** in particular cases.) It is important to note that, in this approach, *conditional* probability, directly defined through the device of called-off bets, is the fundamental notion. Unconditional probability is obtained as a special case when the conditioning event happens to be the certain event--i.e., the whole sample space. This is in contrast to the conventional approach to probability theory, which begins with a definition of unconditional probabilities, and then derives conditional probabilities according to (2.6).

The distribution $\mathbf{w}$ satisfying $\mathbf{w}'\mathbf{A}=0$, whose existence is required for $\mathbf{p}$ to be coherent, need not be unique. Let $W(\mathbf{p})$ denote the closed, convex set consisting of all such $\mathbf{w}$. Given any $\mathbf{w}$ in $W(\mathbf{p})$, the probability measure $\mathbf{P}_\mathbf{w}$ is defined for all subsets of the sample space, not merely the $2n$ events initially considered. This provides a basis for inferences about the possible coherent values for bet prices on further pairs of events which are subsets of the same sample space. Let E_{n+1} and F_{n+1} denote such a further pair of events, and let p_{n+1} denote the bet price for E_{n+1} given F_{n+1}. Then, given that $\mathbf{p}$ is coherent, a necessary and sufficient condition for $(\mathbf{p},p_{n+1})$ to also be coherent is that either $p_{n+1}=\mathbf{P}_\mathbf{w}(E_{n+1}|F_{n+1})$ or else $\mathbf{P}_\mathbf{w}(F_{n+1})=0$ for some $\mathbf{w}$ in $W(\mathbf{p})$. In the latter case, p_{n+1} may coherently assume any value whatever. In the former case, $\mathbf{P}_\mathbf{w}(E_{n+1}|F_{n+1})$ is a continuous, bounded function defined everywhere in $W(\mathbf{p})$, and hence achieves a minimum and maximum (denoted $\hat{p}_{n+1}^-$ and $\hat{p}_{n+1}^+$, respectively) on this set, as well as all values in between. Thus, if $\mathbf{P}_\mathbf{w}(F_{n+1})>0$ for all $\mathbf{w}$ in $W(\mathbf{p})$, then $(\mathbf{p},p_{n+1})$ is coherent if-and-only-if $\hat{p}_{n+1}^- \leqslant p_{n+1} \leqslant \hat{p}_{n+1}^+$. This is de Finetti's "fundamental theorem of probability" (1974, p. 112). In fact, $\hat{p}_{n+1}^-$ and $\hat{p}_{n+1}^+$ are lower and upper conditional probabilities for E_{n+1} given F_{n+1}, indirectly determined by $\mathbf{p}$, in the sense that they represent the lowest selling price and the highest buying price for a unit bet on E_{n+1} given F_{n+1} which would be consistent with $\mathbf{p}$. This notion will be developed further in Section 4.

In order to be *strictly* coherent, a set of bet prices must not only obey the probability laws, but also be consistent with some assignment of positive probability to every outcome in the sample space. In the sense of the preceding discussion, the bookie has implicitly assigned zero probability to outcome j (unconditionally) if that outcome has zero probability under every probability measure which yields the bet prices as conditional probabilities-- i.e., if $w_j=0$ in every non-negative solution to $\mathbf{w}'\mathbf{A}=0$, where $\mathbf{A}$ is the matrix defined above. The implication of Theorem 2 is that there exists a combination of bets for which the bookie will lose positive amounts of money under all those outcomes (and only those outcomes) which he has implicitly assigned zero probability, while winning nothing under the remaining outcomes. Either coherence or strict coherence can be used as the criterion for defining and measuring subjective probability. As Buehler (1976, p. 1057) points out, "the philosophical choice between the two criteria is clearly linked to one's attitude toward the acceptability of subjective probabilities which equal zero." De Finetti (1974) argues that zero probabilities are necessary in order to deal with infinite partitions, and hence favors the weaker criterion, coherence. However, in any physically realizable, which is to say finite, experiment, it appears that strict coherence would be more in accord with ordinary standards of behavior. This issue will be illuminated further by the corresponding distinction between admissible and strictly admissible choices under scoring rules in a later section. Throughout the remainder of this paper, the term "bet prices" will be used to denote conditional or unconditional subjective probabilities elicited under the betting system described above, in order to avoid confusion with probabilities defined, elicited, or derived in other ways.

The conditions under which a set of bet prices is coherent, as given in Theorem 2, have a simple geometric interpretation. A probability distribution $\mathbf{w}$ can be represented as a vector in m-space lying in the standard simplex defined by $\sum_{j=1}^{m} w_j=1$, $\mathbf{w}\geqslant 0$. The bet prices $\mathbf{p}$ are coherent if-and-only-if there exists a distribution $\mathbf{w}$ satisfying the system of equations $\mathbf{w}'\mathbf{A}=0$, which will be referred to as the "bet price constraints." Let $\boldsymbol{\alpha}_i$ denote the i^{th} column vector of $\mathbf{A}$--i.e., the vector whose j^{th} element is $(p_i-E_{ij})F_{ij}$. Then the bet price constraints can be written as $\mathbf{w}'\boldsymbol{\alpha}_i=0$, $i=1, \ldots, n$. Geometrically, $\boldsymbol{\alpha}_i$ is the normal vector of a hyperplane passing through the origin whose intersection with the simplex is the set of all $\mathbf{w}$ which are consistent with the bet price p_i. This is illustrated in Figure 2.1 for the case of a sample space consisting of only three elements, $\Theta=\{\theta_1,\theta_2,\theta_3\}$. Here the simplex of probability distributions is the triangle whose vertices are the unit vectors, each vertex being identified with one element of Θ. The hyperplane normal to the vector $\boldsymbol{\alpha}_i$ pictured is defined by the origin and the two points on the boundary of the simplex labelled A and B. The line segment AB represents the set of probability distributions consistent with p_i, i.e., those $\mathbf{w}$ for which $\mathbf{P}_{\mathbf{w}}(E_i|F_i)=p_i$. Incoherence arises when there is no point in the simplex at which all n bet price hyperplanes intersect.

Some examples of coherence and incoherence for three-element sample spaces are illustrated in Figures 2.2, 2.3, and 2.4. These figures are drawn in the plane defined by $\sum_{j=1}^{m} w_j=1$, in which the simplex appears as an equilateral triangle. If this triangle is scaled so that its height is unity, then the probability distribution (w_1,w_2,w_3) corresponding to any point is determined by letting w_j equal the perpendicular distance from that point to the side opposite the vertex corresponding to θ_j. Where convenient, the notation $p(E|F)$ and $p(E)$ will also be used to denote conditional and unconditional bet prices for the events parenthesized. Figures 2.2a and 2.2b represent the simple case of a complete partition of the sample space, in which $m=n=3$, and $p_i=p(\theta_i)$ for all i. In this case the bet price constraints reduce to $\mathbf{w}=\mathbf{p}$, which, together with the coherence requirement that this must be satisfied for some $\mathbf{w}$ in the simplex, implies $p_1+p_2+p_3=1$. (This is the "total probability theorem.") This condition is satisfied in 2.2a by $\mathbf{p}=(.4,.3,.3)$, and violated in 2.2b by $\mathbf{p}=(.6,.3,.3)$. For each i, the set of $\mathbf{w}$ for which $w_i=p_i$ is a line parallel to the face opposite the vertex corresponding to θ_i. The three lines so determined by $\mathbf{p}$ have a point of mutual intersection (inside the simplex) in 2.2a, whereas in 2.2b they do not. The case of an incomplete partition is illustrated in Figures 2.3a, 2.3b, and 2.3c, where $m=3$, $n=2$, and $p_i=p(\theta_i)$ for $i=1,2$. Here the bet price constraints on $\mathbf{w}$ are $w_1=p_1$ and $w_2=p_2$, which is satisfied by some $\mathbf{w}$ in [the interior of] the simplex only if $p_1+p_2 \leqslant 1$ [<1]. This relation is satisfied in Figures 2.3a and 2.3b by $\mathbf{p}=(.3,.6)$ and $\mathbf{p}=(.4,.6)$, respectively. (Note that the latter choice is not strictly coherent, since it implies $p(\theta_3)=0$, and the corresponding probability distribution is therefore represented by a point on the boundary of the simplex, rather than in the interior.) Figure 2.3c illustrates the choice $\mathbf{p}=(.6,.6)$, which is incoherent. In this case the lines representing the bet price constraints have a point of intersection in the plane, but it is outside the simplex.

Whereas the previous examples illustrate the additive law of probability, the next examples, in Figures 2.4a and 2.4b, illustrate the multiplicative law. For some events E and F, let $p_1=p(EF)$, $p_2=p(F)$, and $p_3=p(E|F)$. The (minimal) relevant sample space is then $\theta_1=EF$, $\theta_2=(1-F)$, $\theta_3=(1-E)F$. Here the bet price constraints are $w_1=p_1$, $w_2=1-p_2$, and $w_1(p_3-1)+w_3p_3=0$, which implies $\frac{w_1}{w_1+w_3}=p_3$ unless $w_1+w_3=0$. Now, coherence requires $w_1+w_2+w_3=1$, whence $p_2=1-w_2=w_1+w_3$, and also $\mathbf{w} \geqslant 0$, so that $p_2 \geqslant p_1 \geqslant 0$--i.e., $p(F) \geqslant p(EF) \geqslant 0$. In particular, $p(F)=0$ implies $p(EF)=0$. Otherwise, if $p_2=p(F)\neq 0$, then the identity $w_1=(w_1+w_3)(\frac{w_1}{w_1+w_3})$ implies $p_1=p_2p_3$--i.e., $p(EF)=p(F)p(E|F)$. (This is the "compound probability theorem.") The first two bet price constraints correspond to lines parallel to the sides opposite the vertices for θ_1 and θ_2, respectively, and the third constraint corresponds to a line passing through the vertex for θ_2 and through a point on its opposite side. In Figure 2.4a, where $\mathbf{p}=(.2,.5,.4)$ satisfies the multiplicative probability law, these three lines intersect at a point in the simplex, whereas in Figure 2.4b, where $\mathbf{p}=(.4,.5,.4)$ violates the multiplicative law, these lines have no point of common intersection.

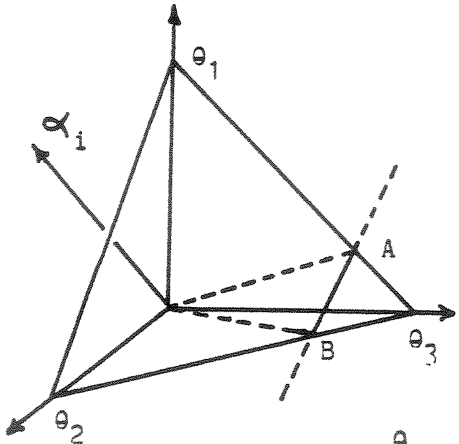


Fig. 2.1 Geometry of a Bet Price Constraint

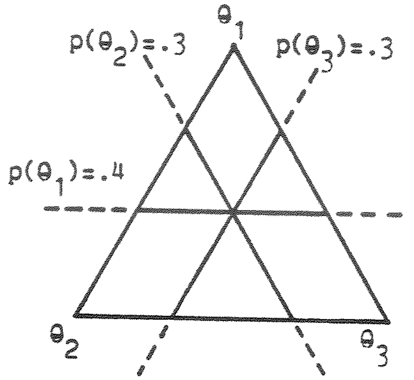


Fig. 2.2a Strict Coherence

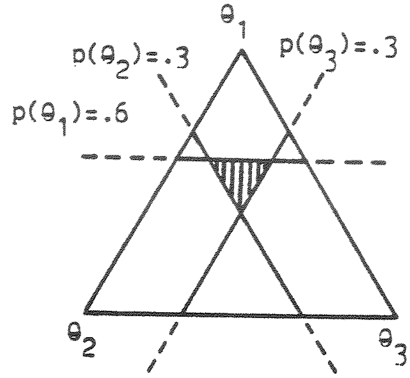


Fig. 2.2b Incoherence

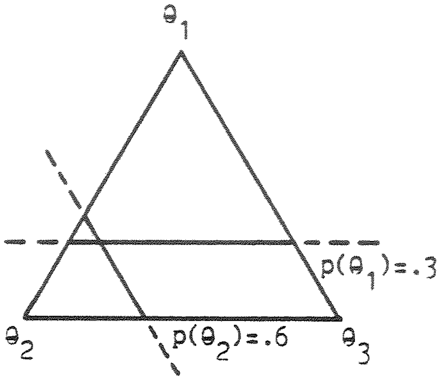


Fig. 2.3a Strict Coherence

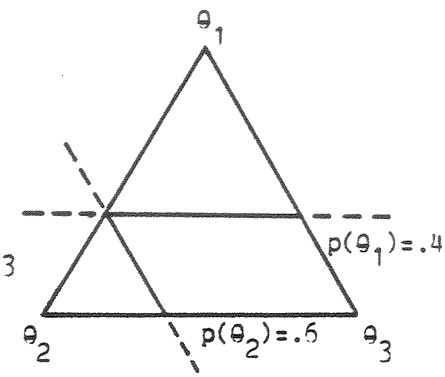


Fig. 2.3b Coherence

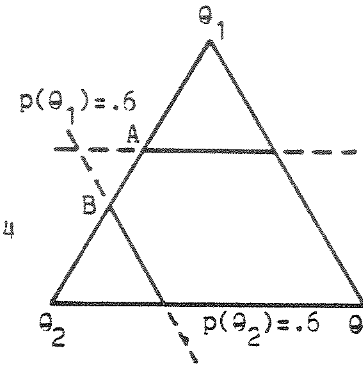


Fig. 2.3c Incoherence

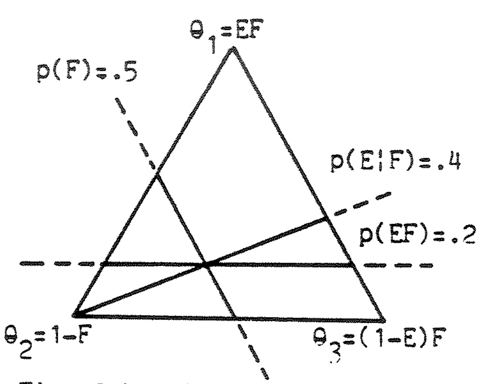


Fig. 2.4a Strict Coherence

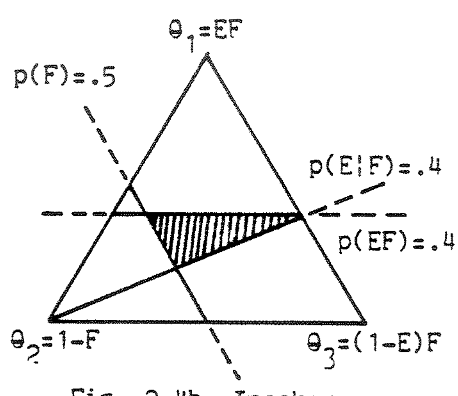


Fig. 2.4b Incoherence

3. Identifying and reconciling incoherence via linear programming

It has been seen that, given a vector **p** of bet prices, the problem of finding a "sure-win" bet (the "bettor's problem") and the alternative problem of finding a corresponding distribution on the sample space (the "bookie's problem") involve finding solutions to systems of linear inequalities. These are textbook linear programming problems, and, moreover, the role of the corollary to Theorem 1 (which is a variant of Farka's Lemma) in the proof of Theorem 2 suggests that the bookie's and bettor's problems are in fact dual to each other. This linear programming application does not appear, however, to have received explicit treatment in the literature, perhaps because the subject of coherence has generally been considered to be of more theoretical than practical interest. The conventional emphasis has instead been on calibration--i.e., obtaining subjective probability assessments which agree with observed frequencies. The seminal paper of Lindley, Tversky, and Brown (1979) has pointed out the relevance of coherence considerations in improving the precision of probability assessments and in combining assessments by different experts, as well as in avoiding mere inconsistency. Their approach to the identification and reconciliation of incoherence is thoroughly Bayesian, and uses a "coherent observer," equipped with a prior distribution and likelihood function, to perform the reconciliation. In this section several geometrically-motivated linear programs for the identification and reconciliation of incoherence will be discussed. It will be seen that, under certain assumptions and conditions, the Bayesian and linear programming approaches closely resemble each other.

As a starting point, consider the following linear program (which will be called LP1), which can be used to distinguish between coherence, strict coherence, and incoherence:

Primal: maximize y_0 (3.1a)

$$\text{subject to } y_0 + y_j + \sum_{i=1}^{n} \alpha_{ij} z_i = 0 \quad j=1, \ldots, m \tag{3.1b}$$

$$\sum_{j=1}^{m} \beta_j y_j = 1 \tag{3.1c}$$

$$y_j \geqslant 0 \quad j=1, \ldots, m\,. \tag{3.1d}$$

Dual: minimize w_0 (3.2a)

$$\text{subject to } \sum_{j=1}^{m} \alpha_{ij} w_j = 0 \quad i=1, \ldots, n \tag{3.2b}$$

$$\sum_{j=1}^{m} w_j = 1 \tag{3.2c}$$

$$w_j + \beta_j w_0 \geq 0 \quad j=1, \ldots, m\,. \tag{3.2d}$$

The parameters of this linear program are the matrix A, whose $(j,i)^{\text{th}}$ element is $\alpha_{ij}=(p_i-E_{ij})F_{ij}$, and a vector $\beta=(\beta_1, \ldots, \beta_m)$ of positive weights. It will be convenient, with no loss of generality, to assume that $\sum_{j=1}^{m}\beta_j=1$ and to consider β to represent the *bettor's* probability distribution on the sample space. In the primal program the natural variables are $z_1, \ldots, z_n$, where z_i represents the number of unit bets on the i^{th} event pair purchased by the bettor. $y_1, \ldots, y_m$ are non-negative slack variables, and y_0 is essentially an artificial variable guaranteeing the existence of a feasible solution. (A starting feasible solution is $y_0=-1$, $y_j=1$ for $j=1, \ldots, m$, and $z_i=0$ for $i=1, \ldots, n$.) The primal program can be interpreted as searching for a combination of bets which maximizes the ratio of the bettor's minimum payoff to his expected payoff, provided the latter quantity can be made positive. Since the quantity $\sum_{i=1}^{n}\alpha_{ij}z_i$ represents the payoff to the bookie under the j^{th} outcome, constraint (3.1b) implies that in every feasible solution y_0+y_j represents the payoff to the bettor under the same outcome. Since y_j is constrained to be non-negative, y_0 is evidently greater than or equal to the bettor's minimum payoff. The bettor's expected payoff is then given by $y_0+\sum_{j=1}^{m}\beta_j y_j$. Star notation ($z_i^*$, y_j^*, w_j^*, etc.) will be used to denote the values of the primal and dual variables in an optimal solution. If all bets have zero expectation for the bettor (a special case of strict coherence) then y_0^* will be -1 in view of constraint (3.1c). In all other cases, when bets with non-zero expectation are possible, y_0^* will be greater than -1, and will represent the minimum payoff (implying $y_j^*=0$ for at least one $j \geq 1$) for some bet with positive expectation. In fact, y_0^* will be the maximum possible minimum payoff among all bets whose expected payoff exceeds the minimum payoff by exactly unity. Moreover, no bet with the same expectation as that of the optimal primal solution can have a larger minimum payoff, since otherwise constraint (3.1c) would not be tight, and such a bet could therefore be scaled up to obtain a feasible solution with a higher objective value, contradicting optimality. Hence, y_0^* must be negative, zero, or positive according to whether the bet prices are strictly coherent, coherent, or incoherent, by the very definitions of these terms. If $y_0^*=0$, then $\mathbf{z}^*=(z_1^*, \ldots, z_n^*)$ is a "can't-lose" bet; and if $y_0^*>0$, then $\mathbf{z}^*$ is a "sure-win" bet. A special case arises when there is a bet for which the payoffs are identical and positive--i.e., the minimum payoff equals the expected payoff--in which case y_0^* is not only positive, but infinite. In this case, the components of the "sure-win" bet can be found in the column of the simplex tableau corresponding to the entering variable which produces an unbounded increase in the objective function. If y_0^* is negative, then the bet $\mathbf{z}^{**}$, defined by $z_i^{**} = \frac{z_i^*}{-y_0^*}$ for all i, achieves the largest possible expected payoff among all bets whose minimum payoff is greater than or equal to -1.

The dual program can be interpreted as searching for a probability distribution on the sample space which is consistent with the bet prices and which is also as close as possible, in a certain sense, to the bettor's distribution. The dual natural variables consist of w_0 together with the elements of the vector $\mathbf{w}=(w_1, \ldots, w_m)$. The solution of the dual program by the simplex algorithm can be visualized in the m-dimensional space in which $\mathbf{w}$ is represented, along the lines of the geometric interpretation of coherence presented in the last section. The feasible region for $\mathbf{w}$ is the set of points in the hyperplane defined by $\sum_{j=1}^{m} w_j=1$ which satisfy the bet price constraints $\mathbf{w}'\mathbf{A}=0$. (Note that the feasible region may contain points outside the simplex, i.e., which do not also satisfy $\mathbf{w} \geqslant 0$.) If this set is non-empty (which is the case if-and-only-if the objective function is bounded in the optimal primal solution) then the solution of the dual program by the primal simplex algorithm involves starting at some feasible point and then moving within the feasible region toward (the interior of) the simplex by maximizing the weighted minimum of the coordinates, with the weights being the reciprocals of the bettor's probabilities. This is seen by rewriting the dual objective (3.2a) as "maximize $-w_0$" and rewriting the constraint (3.2d) as "$-w_0 \leqslant \frac{w_j}{\beta_j}$". If the dual program terminates at a point in the interior of the simplex (i.e. the weighted minimum coordinate is positive, and hence w_0^* is negative), then a positive probability distribution (namely $\mathbf{w}^*$) has been found which agrees with the bet prices, and strict coherence has been established, according to Theorem 2. If termination occurs at a point on the surface of the simplex (i.e., the minimum coordinate is zero, and so is w_0^*), then a semi-positive distribution has been found, establishing coherence but not strict coherence. Finally, if termination occurs outside the simplex (i.e., one of the final coordinates is negative, hence w_0^* is positive), then no appropriate semi-positive distribution exists, and incoherence is established. Of course, by the duality theorem of linear programming, $w_0^*=y_0^*$, so that LP1 may be taken as a constructive proof of Theorem 2. The special case in which $w_0^*=y_0^*=-1$, which arose in the primal program when every bet had zero expectation for the bettor, is seen in the dual program to represent the case in which the *bettor's* probability distribution is consistent with the bet prices-- i.e., the optimal dual solution is $w_j^*=\beta_j$, $j=1, \ldots, m$.

If a set of bet prices is found to be incoherent (e.g., by solving LP1), then presumably it will be desired to revise them so as to reconcile the incoherence. Properly considered, this ought to involve introspection and careful reassessment on the part of the bookie. However, insofar as the constraints imposed by coherence may be too numerous or subtle to keep in mind during this process, it might be useful or even necessary to have an external procedure for identifying coherent sets of bet prices which are in some sense "close" to the original incoherent set, in order to help the bookie explore his alternatives. A Bayesian reconciliation scheme has been presented by Lindley, Tversky, and Brown (1979). In their "internal approach," a coherent observer is introduced who considers the bookie's true, coherent bet prices as uncertain parameters to be estimated. The observer has a (continuous) prior distribution for these parameters and a likelihood function specifying the distribution of the errors in the bookie's stated bet prices given his true bet prices. The posterior distribution is then computed using Bayes' Theorem, and a vector of revised, coherent prices can be obtained as the posterior expected value of the true bet price vector, subject to the constraints imposed by coherence. If the coherence constraints are non-linear, then the set of possible coherent vectors may be non-convex, in which case the posterior mode, rather than the posterior expected value, must be used, since the barycenter of a distribution of mass on a non-convex set may be a point outside the set. Unfortunately, there are several practical difficulties associated with the strict Bayesian approach, namely the assessment of the "core" distributions required in the conditioning process, and the complexity of the calculations.

To simplify matters, Lindley *et al* suggest a least-squares approach to finding a reconciled bet price vector, which is consistent with the assumption of a flat prior distribution and normally distributed errors. If the errors are also assumed independent, the reconciliation is obtained by minimizing the weighted sum of squares:

$$\sum_{i=1}^{n} \tau_i^2 (p_i - \pi_i)^2 \tag{3.3}$$

over the set of all coherent π, where $\tau = (\tau_1, \ldots, \tau_n)$ is a vector of positive weights equal to the reciprocals of the error standard deviations. Since the elements of $\mathbf{p}$ are restricted to the unit interval, they can at best only be approximately normally distributed with respect to the "true" bet prices. Therefore, it may be appropriate to assume that some transform of each bet price, say $F(p_i)$, is normally distributed, and then minimize the sum of the squared differences of the transforms:

$$\sum_{i=1}^{n} \tau_i^2 (F(p_i) - F(\pi_i))^2 \; . \tag{3.4}$$

In particular, the log odds transform is recommended: $F(p) = log(\frac{1-p}{p})$. Lindley *et al* refer to this transformation as the "choice of metric"-- probability metric, log-odds metric, etc.--and suggest that the choice of metric should reflect the transformation under which the error variance is most nearly constant. For purposes of later comparison, note that the sum of squares (3.3) is the squared distance between $\mathbf{p}$ and π, following a linear transformation by the matrix $diag(\tau_1, \ldots, \tau_n)$, using the l_2 norm. An alternative choice of metric would be to minimize the distance between these vectors using a different norm. For example, in the l_1 norm the corresponding distance is

$$\sum_{i=1}^{n} \tau_i |p_i - \pi_i| \tag{3.5}$$

and in the l_∞ norm the distance is

$$\max_i \tau_i |p_i - \pi_i| \, . \tag{3.6}$$

Of course, a vector minimizing one of these distances would not exactly correspond to the mode of a normal posterior distribution, and the weights τ would not be interpretable as reciprocal variances, but simply as a set of (subjectively chosen) confidence or precision factors.

The various minimizations suggested above must all be performed subject to the coherence constraints on π, which consist of a set of equalities and inequalities of the form

$$h_k(\pi) = 0 \; [\, \geqslant 0 \,], \; k=1, \ldots, K \tag{3.7}$$

where the functions $\{h_1, \ldots, h_K\}$ consist of sums and products of the elements of π and constants, representing the requirements of the additive, multiplicative, and convexity laws of probability. They are essentially implicit functions determined by the equations $\sum_j (\pi_i - E_{ij}) F_{ij} w_j = 0$, $i=1, \ldots, n$, and $\sum_j w_j = 1$, together with the inequalities $w_j \geqslant 0$, $j=1, \ldots, m$. A practical problem may arise if the multiplicative probability law is involved, in which case some of the constraint functions will be nonlinear. The resulting constraint set may be non-convex, and exact global minimization by systematic nonlinear programming methods may therefore be difficult. Moreover, if the constraint set is pathologically shaped, the implicit assumption of a flat distribution on it may be questionable.

An alternative approach, suggested by the geometric representation of incoherence emphasized in this paper, would be to represent the reconciled assessment in terms of a corresponding distribution on the sample space--i.e., $\pi_i = \mathbf{P}_\mathbf{w}(E_i | F_i)$, $i=1, \ldots, n$-- and then perform the minimization over $\mathbf{w}$. The constraint set for $\mathbf{w}$ is in all cases a convex set, namely the standard simplex in m-space. The practical difficulties with this approach are associated with the nature of the resulting objective function, since

$$p_i - \mathbf{P}_\mathbf{w}(E_i | F_i) = \frac{\sum_j (p_i - E_{ij}) F_{ij} w_j}{\sum_j F_{ij} w_j} = \frac{\sum_j (p_i - E_{ij}) F_{ij} w_j}{\mathbf{P}_\mathbf{w}(F_i)} . \tag{3.8}$$

Note the presence in the denominator of the term $\mathbf{P_w}(F_i)$. Unless the probabilities being assessed are all unconditional (in which case $F_i=\Theta$ and $\mathbf{P_w}(F_i)=1$ for all i), the quantity (3.8) is a nonlinear and not necessarily convex function of $\mathbf{w}$. Thus, unfortunately, in changing variables from $\boldsymbol{\pi}$ to $\mathbf{w}$ in order to obtain a convex constraint set for performing the minimizations (3.3), (3.5), or (3.6), a non-convex objective function may be obtained, which again may be difficult to minimize globally. One way around this difficulty is to simply ignore the term in the denominator of (3.8), and concentrate on minimizing an appropriate function of the quantities

$$\mathbf{w}'\boldsymbol{\alpha}_i = \sum_j (p_i - E_{ij}) F_{ij} w_j, \; i=1, \ldots, n, \tag{3.9}$$

where $\boldsymbol{\alpha}_i$ again denotes the i^{th} column vector of the matrix A. This seemingly *ad hoc* linearization of the objective function has a significant and interesting geometric interpretation in the space of probability distributions on the sample space, for the quantity $\mathbf{w}'\boldsymbol{\alpha}_i$ is proportional to the Euclidean distance from $\mathbf{w}$ to the nearest point in the hyperplane of the simplex which satisfies the i^{th} bet price constraint. To show this, let $\mathbf{u}_i(\mathbf{w})$ denote the vector which minimizes $||\mathbf{w}-\mathbf{u}||$ subject to $\mathbf{u}'\boldsymbol{\alpha}_i=0$ and $\sum_j u_j=1$, and let $d_i(\mathbf{w})=||\mathbf{w}-\mathbf{u}_i(\mathbf{w})||$. Note that if $\mathbf{u}_i(\mathbf{w}) \geqslant 0$, then $\mathbf{u}_i(\mathbf{w})$ is the closest distribution to $\mathbf{w}$ (in the Euclidean sense) which yields p_i as the conditional probability for E_i given F_i. Necessarily, $||\mathbf{w}-\mathbf{u}_i(\mathbf{w})||$ is proportional to the the vector $\boldsymbol{\alpha}_i^o$, whose j^{th} component is $\alpha_{ij}^o = \alpha_{ij} - \frac{1}{m}\sum_j \alpha_{ij}$, which is obtained by projecting $\boldsymbol{\alpha}_i$ on the hyperplane $\sum_j w_j=0$. From the definition of $d_i(\mathbf{w})$, it follows that $\mathbf{u}_i(\mathbf{w})=\mathbf{w}\pm\frac{d_i(\mathbf{w})}{||\boldsymbol{\alpha}_i^o||}\boldsymbol{\alpha}_i^o$. Enforcing $\mathbf{u}_i(\mathbf{w})'\boldsymbol{\alpha}_i=0$, and noting that $\boldsymbol{\alpha}_i^{o\prime}\boldsymbol{\alpha}_i^o=\boldsymbol{\alpha}_i'\boldsymbol{\alpha}_i^o$, yields:

$$d_i(\mathbf{w}) = \frac{|\mathbf{w}'\boldsymbol{\alpha}_i|}{||\boldsymbol{\alpha}_i^o||} = \frac{|\mathbf{w}'\boldsymbol{\alpha}_i|}{\sqrt{\sum_j \alpha_{ij}^2 - \frac{1}{m}(\sum_j \alpha_{ij})^2}} \; . \tag{3.10}$$

Note that $\mathbf{w}'\boldsymbol{\alpha}_i$ is the expected payoff to the bookie for a unit bet placed on E_i given F_i, under the distribution $\mathbf{w}$. On the other hand, it is evident from the expansion on the RHS of (3.10) that the quantity $||\boldsymbol{\alpha}_i^o||$, which is a sort of normalizing factor, is $\sqrt{m}$ times the standard deviation of the payoff for a unit bet on E_i given F_i under the *uniform* distribution. It is readily shown that, if $F_{ij}=1$ for all j (i.e., if p_i is an unconditional bet price for E_i), then $||\boldsymbol{\alpha}_i^o||$

is independent of p_i. If, however, $F_{ij}=0$ for at least one j, then $||\alpha_i^o||^2$ is a convex quadratic function of p_i which is minimized when p_i equals the ratio of the number of outcomes in Θ for which $E_iF_i=1$ to the number of outcomes for which $F_i=1$--i.e., the value for the conditional probability of E_i given F_i which is obtained under the uniform distribution.

In the space in which distributions on Θ are represented, the distance $d_i(\mathbf{w})$ appears to be reasonable measure of the "error" in the bet price p_i when the "true" distribution is $\mathbf{w}$. Minimization on the simplex of an appropriate convex function of these distances will yield a distribution which satisfies an heuristic admissibility criterion, namely that no other distribution exists which is uniformly closer to all the hyperplanes determined by the bet price constraints. In the examples of incoherence illustrated in the previous section, the sets of distributions which are admissible in this sense are represented by the shaded areas in Figures 2.2b and 2.4b, and the line segment AB in Figure 2.3c. Let $\gamma=(\gamma_1, \ldots, \gamma_n)$ be vector of positive weights representing relative confidence or precision under this measure of "bet price error," incorporating the normalizing factors $||\alpha_i^o||$, $i=1, \ldots, n$, suggested by (3.10). Then, by analogy with (3.5), (3.6), and (3.3), some possible objective functions for minimization are:

$$\sum_{i=1}^{n} \gamma_i |\mathbf{w}'\alpha_i| \tag{3.11}$$

or

$$\max_i \gamma_i |\mathbf{w}'\alpha_i| \tag{3.12}$$

or else the quadratic form

$$\sum_{i=1}^{n} \gamma_i^2 (\mathbf{w}'\alpha_i)^2 = \mathbf{w}'(\mathbf{AMA}')\mathbf{w} \tag{3.13}$$

where $\mathbf{M}=diag(\gamma_1^2, \ldots, \gamma_n^2)$. (More generally, $\mathbf{M}$ could be any positive definite matrix.) The minimization on the simplex of either (3.11) or (3.12) is a straightforward linear program, and (3.13) is a quadratic program with linear constraints. However, (3.12) appears to be a much more suitable objective function for practical application than (3.11). The close relation between (3.11) and the constraint $\sum_j w_j=1$ suggests that the solution may be highly sensitive to relatively small changes in the weights, and will tend to be an extreme point of the admissible region. (In particular, for the case of an incoherent partition, only the bet price with the largest weight is likely to be revised.) By comparison, it appears that minimization of (3.12) will generally yield an interior point of the admissible set--in fact, with an appropriate choice of weights, any admissible point can be reached--and it will be seen to yield essentially the same solution as the quadratic minimization (3.13) in the case of a partition.

The linear program representing the minimization of (3.12) will now be discussed in some detail. First, note that minimizing $\{\max_i \gamma_i |\mathbf{w}'\boldsymbol{\alpha}_i|\}$ is equivalent to finding the smallest number v such that

$$\frac{-v}{\gamma_i} \leqslant \mathbf{w}'\boldsymbol{\alpha}_i \leqslant \frac{v}{\gamma_i},\ i=1,\ldots,n \tag{3.14}$$

for some $\mathbf{w}$ in the simplex. For added flexibility, let each weight γ_i be replaced by a pair of possibly-unequal positive weights γ_i^+ and γ_i^-, with γ_i^+ substituted for γ_i on the left and γ_i^- substituted for γ_i on the right in (3.14). This allows for the possibility that, in seeking an optimal reconciled bet price vector, positive and negative deviations from each of the initial bet prices will be weighted differently, which might be especially desirable for bet prices very near to 0 or 1. The corresponding primal/dual pair of linear programs is then:

Primal: maximize y_0 (3.15a)

$$\text{subject to } y_0 + \sum_{i=1}^{n} \alpha_{ij}(z_i^+ - z_i^-) \leqslant 0 \tag{3.15b}$$

$$\sum_{i=1}^{n} \left(\frac{z_i^+}{\gamma_i^+} + \frac{z_i^-}{\gamma_i^-}\right) = 1 \tag{3.15c}$$

$$z_i^+ \geqslant 0,\ z_i^- \geqslant 0,\ i=1,\ldots,n. \tag{3.15d}$$

Dual: minimize v (3.16a)

$$\text{subject to } \frac{-v}{\gamma_i^+} \leqslant \sum_{j=1}^{m} \alpha_{ij} w_j \leqslant \frac{v}{\gamma_i^-} \tag{3.16b}$$

$$\sum_{j=1}^{m} w_j = 1 \tag{3.16c}$$

$$w_j \geqslant 0,\ j=1,\ldots,m. \tag{3.16d}$$

Note that in the corresponding primal program the unrestricted-sign variable z_i of LP1 (the number of unit bets on E_i given F_i) has been replaced by a pair of non-negative variables, z_i^+

and z_i^-, representing its positive and negative parts. The primal objective is still the maximization of the bettor's minimum payoff, but the external constraint is now a prescribed value for the weighted sum of the numbers of unit bets bought and sold, rather than a prescribed value for the difference between the minimum payoff and the expected payoff. In view of the dual constraints (3.16b), the optimal objective value can never be negative. In the primal program this is reflected in the fact that the bettor can always satisfy constraint (3.15c) by buying and selling equal numbers of bets on any event, which is equivalent to not betting, since the buying and selling prices are equal.

An interesting special case is obtained by letting $\gamma_i^+ = \frac{1}{(1-p_i)}$ and $\gamma_i^- = \frac{1}{p_i}$ for every i, whence (3.15c) becomes

$$\sum_{i=1}^{m}((1-p_i)z_i^+ + p_i z_i^-) = 1. \tag{3.17}$$

Note that the quantity on the left is the bettor's *a priori* maximum payoff--that is, the amount he would receive from the bookie if he won every bet. This is "prior" to an analysis of the logical dependencies among the events, which might show certain joint outcomes of the events to be impossible, in which case certain combinations of bets could not be won. This constraint may be considered to describe the situation in which the bookie has finite resources, and will only accept bets up to the amount he can "cover" by separately matching each bet with the amount the bettor might win. This weighting also has an interesting interpretation in the dual program. Recall that

$$p_i - \mathbf{P}_{\mathbf{w}}(E_i|F_i) = \frac{\mathbf{w}'\boldsymbol{\alpha}_i}{\mathbf{P}_{\mathbf{w}}(F_i)}, \tag{3.18}$$

whence the differences between the initial bet prices and the conditional probabilities based on the distribution $\mathbf{w}$ have the same signs as, and are approximately proportional to, the quantities $\mathbf{w}'\boldsymbol{\alpha}_i$, whose weighted deviations from zero are minimized in the dual program. Under this weighting, a positive deviation from zero of $\mathbf{w}'\boldsymbol{\alpha}_i$, corresponding to a positive difference between p_i and $\mathbf{P}_{\mathbf{w}}(E_i|F_i)$, is weighted in proportion to $\frac{1}{p_i}$ --i.e., in inverse proportion to the maximum positive difference possible (which is obtained when $\mathbf{P}_w(E_i|F_i)=0$). A corresponding effect is obtained with respect to negative deviations. Roughly speaking, initial bet prices near zero tend to be revised upward rather than downward in the reconciliation process using these weights, and vice versa for initial bet prices near unity. The reconciliation process, in this case, tends to pull all the elements of $\mathbf{p}$ toward the value $\frac{1}{2}$, insofar as this can be done coherently.

The reconciliation scheme of LP2 resembles a Bayesian approach formulated in the m-space of probability distributions on the sample space rather than the n-space of bet price vectors. In fact, for the simple case in which the events constitute a partition, where every coherent bet price vector is also a distribution on the sample space, LP2 (with appropriate weights) yields the same reconciled values as the "internal approach" of Lindley *et al* under the "probability metric"--i.e., using the quadratic minimization (3.3), which is also the same as (3.13) in this case. Here the dual objective (maximum weighted deviation) in LP2 and the weighted sum of squared deviations in (3.3) are both minimized when their respective weighted deviations are all equal--i.e., when the deviations are proportional to the inverses of the corresponding weights. Let $\gamma_i^+ = \gamma_i^- = \tau_i^2$ for all i. Then, letting $\boldsymbol{\pi}^*$ denote the coherent bet price vector which minimizes (3.3) and letting $\mathbf{P}_w^*$ denote the probability measure corresponding to the optimal solution to LP2, we have:

$$\pi_i^* = \mathbf{P}_w^*(E_i|F_i) = p_i + \frac{1-\sum_{j=1}^{n} p_j}{\tau_i^2 \sum_{j=1}^{n} \tau_i^{-2}}, \quad i=1, \ldots, n. \tag{3.20}$$

To illustrate the application of LP2 to an actual problem, more difficult than a simple partition, consider the following example of an incoherent assessment which was given in Lindley *et al:* $p(H)=.33$, $p(C)=.27$, $p(D)=.23$, $p(\bar{N})=.12$, $p(H|N)=.41$, $p(C|N)=.31$, and $p(D|N)=.28$. Here $m=4$, $n=7$, and H, C, D, and $\bar{N}$ form a partition of the sample space, so that H, C, and D also form a partition of N. Both the additive law and the multiplicative law are violated--e.g., $p(H)+p(C)+p(D)+p(\bar{N})\neq 1$, $p(H)\neq p(H|N)(1-p(\bar{N}))$, etc.. Four iterations of the simplex algorithm on LP2, with all weights equal (for lack of further information) yields the following reconciled values: $\mathbf{P}^*_w(H)$=.3428, $\mathbf{P}^*_w(C)$=.2816, $\mathbf{P}^*_w(D)$=.2428, $\mathbf{P}^*_w(\bar{N})$=.1328, $\mathbf{P}^*_w(H|N)$=.3953, $\mathbf{P}^*_w(C|N)$=.3247, and $\mathbf{P}^*_w(D|N)$=.28.

The features of LP1 and LP2 can be combined in a single program by incorporating into LP1 the primal constraint (3.15c) from LP2 in Lagrange form, using a multiplier λ. The primal objective then becomes:

$$\text{maximize } y_0 - \lambda \sum_{i=1}^{n} \left(\frac{z_i^+}{\gamma_i^+} + \frac{z_i^-}{\gamma_i^-} \right) \tag{3.21}$$

and the bet price constraints in the dual program become:

$$-\lambda \gamma_i^+ \leqslant \sum_{j=1}^{m} \alpha_{ij} w_j \leqslant \lambda \gamma_i^-, \; i=1, \ldots, n. \tag{3.22}$$

Here the primal program describes the situation in which, from every bet, the bookie is taking a "cut" which is proportional to λ, and also inversely proportional to the corresponding weight (γ_i^+ or γ_i^-). That is, the bookie takes relatively larger cuts from those bets for which his bet prices have low confidence factors. In the dual program λ represents an overall factor by which all of the bet price constraints have been relaxed. The behavior of the optimal solution can be investigated as a function of λ by parametric programming. For any given value of λ, the sign of the optimal objective value plays the same role as in LP1 in determining whether a probability distribution consistent with the (relaxed) constraints has been found, or whether a "sure-win" bet (taking into account the bookie's cut) has been found. The minimum value of λ for which a probability distribution exists (i.e., the smallest overall cut for which no "sure-win" bet exists), is $\lambda = \nu^*$, the optimal objective value that would be obtained in LP2 using the same weights and bet prices. For any $\lambda > \nu^*$, the optimal solution will be affected by both β and γ, and the reconciled bet prices thus determined will generally differ at least slightly from the original bet prices, even if the original assessment was coherent. In this case the elements of β are analogous to the parameters of a prior distribution on the simplex in a Bayesian model, and the elements of γ are analogous to parameters of a likelihood function. Of course, the linear program should not be applied as if it arose from a true Bayesian model--that is, subjectively assessing β and γ once-and-for-all, and accepting the resulting solution. Instead, it appears suitable for use in an interactive process in which the bookie could explore his "admissible frontier" of coherent alternatives, adjusting the parameters until satisfied with the solution. The parameterization of the linear programs described here is simple enough for illustrative purposes but also appears flexible enough for practical application. (Many other parameterizations are possible, of course.) Although β and γ do not correspond exactly to parameters of prior distributions or likelihood functions, they are nonetheless readily interpretable in terms of their effects in steering the optimal solution toward a specified "prior" distribution and/or yielding a reconciliation in which the original bet prices with the highest "confidence" or precision are revised the least.

4. Lower and upper probabilities for the unfair bookie

As noted earlier, in situations involving many events and subtle interdependencies, it may be difficult if not impossible for the bookie to keep in mind all the constraints of coherence while attempting to articulate a set of bet prices which he judges "fair." In such cases he must either derive his bet prices from a previously assessed probability distribution on the sample space, or else obtain the help of an external agent to determine whether his initial subjective bet prices are coherent and to explore nearby coherent alternatives. Therefore, it may be ambiguous to define a person's subjective probabilities for a set of events as his introspectively-obtained coherent bet prices without also specifying by what means coherence is to be verified and incoherence reconciled, if necessary. The acknowledgement that an incoherent initial assessment is possible not only implies the need for procedures for identifying and reconciling incoherence, but also casts some doubt on the validity of initial assessments which are coherent, since they may be coherent only fortuitously. This suggests that the elicitation procedure should be extended in order to obtain additional information which could be used to revise or adjust the initial assessment regardless of whether it is incoherent--in particular, information concerning the relative precision or confidence attached to each of the original bet prices. Enforcement of the coherence constraints would then provide a basis for jointly improving the precision of the separate bet prices. This could provide an important practical tool for improving probability assessments for certain "target" events, by enabling avai lable subjective information concerning other, related events to be brought to bear in a systematic way. The Bayesian approach to this problem is to introduce a hierarchy of probabilities--i.e., probability distributions on probabilities. By restricting the posterior joint distribution of the true values to the set of coherent possibilities, an improvement in precision is manifested in the fact that the variances of the posterior marginal distributions of the separate probabilities will generally be less than the error variances of the initial assessment, even if a flat prior is assumed. However, the parameters or hyperparameters whose values must be elicited to describe these distributions may be difficult to interpret subjectively. In the simplified least-squares reconciliation method of Lindley, Tversky, and Brown (1979), only two numbers need to be elicited for each event, namely the initial assessment of the probability and the variance of its error with respect to the true value. An example is presented in which the variances were elicited by asking the subject to state a range of plausible values for each probability, after which "the quoted ranges were interpreted as multiples of standard deviations..." This section will discuss a conceptually and operationally simpler method for eliciting and utilizing information concerning the precision of subjective probabilities, in which the bookie is asked to specify his uncertainty about one event conditional on another in terms of two numbers which are interpreted as his buying and selling prices for unit bets, when the two prices are not required to be equal. This is the notion of "lower and upper probabilities," which was axiomatized by Koopman (1940) and given a betting interpretation by Smith (1961). (A controversial statistical model was also presented by Dempster (1968).) Lower and upper probabilities have not been highly popular in practice, even among Bayesians (see, e.g., the discussions to Smith (1961) and Dempster (1968)), partly because they are not as easily manipulated as the parameters of hierarchical models by conventional analytical techniques. It will be seen, however, that they provide a basis for a natural generalization of the coherence theorem of Section 2, and are readily incorporated into linear programming models for improving precision and reconciling incoherence.

For the same n pairs of events and same sample space considered throughout this paper, let the bookie announce his *buying* price, p_i^-, and his *selling* price, p_i^+, for a unit bet on E_i conditional on F_i, for every i. The bettor then places his bets by choosing a *non-negative* $2n$-vector $(\mathbf{z}^+,\mathbf{z}^-)$, where z_i^+ is the number of unit bets on E_i given F_i which he wishes to buy (at price p_i^+), and z_i^- is the number he wishes to sell (at price p_i^-). That is, the bettor must buy at the bookie's selling price, and vice versa. The net gain to the bookie for the i^{th} event pair will then be equal to $((p_i^+-E_i)z_i^+-(p_i^--E_i)z_i^-)F_i$, and his total net gain under the j^{th} outcome in the sample space will be

$$t_j(\mathbf{z}^+,\mathbf{z}^-;\mathbf{p}^+,\mathbf{p}^-) = \sum_{i=1}^{n}((p_i^+-E_{ij})z_i^+-(p_i^--E_{ij})z_i^-)F_{ij} \ . \tag{4.1}$$

The bettor is free to both buy and sell bets on the same event, although it will not be profitable for him to do so if $p_i^+ > p_i^-$. Let the payoff vector, $\mathbf{t}(\mathbf{z}^+,\mathbf{z}^-;\mathbf{p}^+,\mathbf{p}^-)$, be defined as the m-vector whose j^{th} element is $t_j(\mathbf{z}^+,\mathbf{z}^-;\mathbf{p}^+,\mathbf{p}^-)$. Let the bet prices be defined, as before, to be [strictly] coherent if-and-only-if there does not exist a combination of bets for which the payoff vector is [semi-] negative. Then the following generalized version of Theorem 2 is obtained:

THEOREM 2': The buying/selling bet prices $(\mathbf{p}^-,\mathbf{p}^+)$ are [strictly] coherent if-and-only-if there exists a [positive] probability distribution $\mathbf{w}$ on Θ, and a corresponding probability measure $\mathbf{P_w}$ on all subsets of Θ, such that for every i, either $p_i^- \leqslant \mathbf{P_w}(E_i|F_i) \leqslant p_i^+$, or else $\mathbf{P_w}(F_i)=0$.

Proof: Note that the payoff vector is given by

$$\mathbf{t}(\mathbf{z}^+,\mathbf{z}^-;\mathbf{p}^+,\mathbf{p}^-) = [\mathbf{A}^+|-\mathbf{A}^-](\mathbf{z}^+,\mathbf{z}^-) = \mathbf{A}^+\mathbf{z}^+ - \mathbf{A}^-\mathbf{z}^- \tag{4.2}$$

where $\mathbf{A}^+$ and $\mathbf{A}^-$ are the $m \times n$ matrices whose $(j,i)^{th}$ elements are $\alpha_{ij}^+=(p_i^+-E_{ij})F_{ij}$ and $\alpha_{ij}^-=(p_i^--E_{ij})F_{ij}$, respectively. By applying Theorem 1 to the matrix $[\mathbf{A}^+|\mathbf{A}^-]$ it follows that either there exists a bet vector for which the corresponding payoff vector is [semi-] negative, or else there exists a semi-positive [positive] vector $\mathbf{w}$ satisfying $\mathbf{w}'[\mathbf{A}^+|\mathbf{A}^-] \geqq 0$. Expanding this vector inequality into n pairs of scalar inequalities, and defining the probability measure as in the proof of Theorem 2, completes the proof.

A similar result is proved, somewhat less transparently, by Smith (1961), in terms of odds for bets on one event "against" another. Based on this theorem, it can be shown that coherent buying and selling bet prices obey the laws of lower and upper probabilities given as axioms by Koopman (1940), in essentially the same way that coherent fair bet prices were shown to obey the additive and multiplicative laws in Section 2. The bookie's buying and selling prices may be considered to provide partial information, in the form of lower and upper bounds, on his fair bet prices. Having stated a willingness to buy at the price p_i^-, he would presumably also buy at a lower price (if possible), and he might even buy at a higher price (if necessary), but he could not simultaneously *sell* at any lower price than p_i^- without inviting certain loss. Similarly, his initially stated selling price, p_i^+, represents an upper bound on his maximum buying price.

Thus, his fair bet prices presumably satisfy $p_i^- \leqslant p_i \leqslant p_i^+$ for all i. Theorem 2' states that his buying and selling prices are coherent if-and-only-if there exists such a set of coherent fair bet prices. The latter quantities are called "medial odds" by Smith (1961).

The generality of different buying and selling prices can be incorporated into the linear programming models of the last section by a trivial modification in which each parameter α_{ij} in the constraints of the original primal problem is replaced by the pair of parameters α_{ij}^+ and α_{ij}^- defined in the proof above, which are associated with the positive and negative parts of z_i, respectively. For example, in LP1, the primal constraint (3.1b) would be replaced by:

$$y_0 + y_j + \sum_{i=1}^{n} (\alpha_{ij}^+ z_i^+ - \alpha_{ij}^- z_i^-) = 0,\ j=1, \ldots, m, \tag{4.3a}$$

$$z_i^+ \geqslant 0,\ z_i^- \geqslant 0,\ i=1, \ldots, n. \tag{4.3b}$$

The dual constraint (3.2b) would correspondingly be replaced by:

$$\sum_{j=1}^{m} \alpha_{ij}^+ w_j \geqslant 0,\ \ i=1, \ldots, n, \tag{4.4a}$$

$$\sum_{j=1}^{m} \alpha_{ij}^- w_j \leqslant 0,\ \ i=1, \ldots, n. \tag{4.4b}$$

In the geometric interpretation of the dual program, for each i, p_i^- and p_i^+ determine the orientations of a pair of hyperplanes in m-space which pass through the origin and are normal to the vectors $\boldsymbol{\alpha}_i^-$ and $\boldsymbol{\alpha}_i^+$, which are the i^{th} column vectors of the matrices $\mathbf{A}^-$ and $\mathbf{A}^+$, respectively. The set of points in the simplex lying on or "below" the first hyperplane (i.e., satisfying $\mathbf{w}'\boldsymbol{\alpha}_i^- \leqslant 0$) and on or "above" the second (i.e., satisfying $\mathbf{w}'\boldsymbol{\alpha}_i^+ \geqslant 0$) is the set of distributions $\mathbf{w}$ for which $p_i^- \leqslant \mathbf{P_w}(E_i|F_i) \leqslant p_i^+$, or else $\mathbf{P_w}(F_i)=0$. The set of buying and selling prices is [strictly] coherent if-and-only-if the intersection of all n such sets, denoted $W(\mathbf{p}^-,\mathbf{p}^+)$, is non-empty [contains an interior point of the simplex]. If, for some i, F_i is not the certain event, then the intersection of the two hyperplanes determined by p_i^- and p_i^+ contains all those points on the boundary of the simplex for which $\mathbf{P_w}(F_i)=0$. If coherence or strict coherence is established by solving this linear program, then the optimal dual solution yields a distribution $\mathbf{w}^*$ which determines a set of coherent fair bet prices, namely $\pi_i^*=\mathbf{P}_{\mathbf{w}}^*(E_i|F_i)$, $i=1, \ldots, n$, lying between (or equalling one of) the respective buying and selling prices. In particular, $\mathbf{w}^*$ has the property that it is the closest such distribution to the "bettor's distribution," $\boldsymbol{\beta}$, in the sense discussed earlier.

A similar modification of LP2 can be used to reconcile as well as identify incoherence. The primal constraint (3.15b) is modified as in (4.3a), and the dual constraint (3.16b) is replaced by:

$$\sum_{j=1}^{m} \alpha_{ij}^{+} w_j \geqslant \frac{-v}{\gamma_i^{+}} \tag{4.5a}$$

$$\sum_{j=1}^{m} \alpha_{ij}^{-} w_j \leqslant \frac{v}{\gamma_i^{-}} \tag{4.5b}$$

where γ_i^- is now interpreted as a precision factor for p_i^- and γ_i^+ is the corresponding precision factor for p_i^+. Note that if $(\mathbf{p}^-,\mathbf{p}^+)$ is coherent and if $p_i^- < p_i^+$ for all i, the optimal objective value may be negative. In the primal problem the interpretation is that the constraint (3.15c) may force the bettor to make a combination of bets which will lose money for him under some outcomes, since he no longer has the option of not betting (i.e., he can no longer buy and sell at the same price). In the dual problem, the interpretation is that a distribution may exist which satisfies all the original bet price constraints (4.4a,b) with strict inequality--i.e., the set $W(\mathbf{p}^-,\mathbf{p}^+)$ has an interior point.

If $(\mathbf{p}^-,\mathbf{p}^+)$ is coherent, then a joint improvement in the precision (in the sense of a narrowing of the intervals $[p_i^-,p_i^+]$, $i=1, \ldots, n$) may be obtained for the same reason that a coherent assignment of fair bet prices may place non-trivial upper and lower bounds on the possible coherent values for a fair bet price on some further, related event. That is, the set $W(\mathbf{p}^-,\mathbf{p}^+)$ determines upper and lower bounds on fair bet prices for all event pairs which are subsets of the same sample space, which, in the case of the event pairs originally considered, may be tighter bounds than the stated buying and selling prices. (Recall that the stated buying price is interpretable as a lower bound on the bookie's minimum selling price; it may not be the greatest lower bound implied by his overall assessment.) The improved lower and upper bounds, denoted $[\hat{p}_i^-,\hat{p}_i^+]$, $i=1, \ldots, n$, can accordingly be defined as:

$$\hat{p}_i^- = \min_{\mathbf{w}} \mathbf{P}_{\mathbf{w}}(E_i \mid F_i) \tag{4.6a}$$

$$\hat{p}_i^+ = \max_{\mathbf{w}} \mathbf{P}_{\mathbf{w}}(E_i | F_i) \tag{4.6b}$$

where the minimization and maximization are with respect to all $\mathbf{w}$ in $W(\mathbf{p}^-, \mathbf{p}^+)$. Necessarily, $p_i^- \leqslant \hat{p}_i^- \leqslant \hat{p}_i^+ \leqslant p_i^+$ for all i. The simplest example of this is the two-fold partition, for which $\hat{p}^+(E) = \min\{p^+(E), 1-p^-(\bar{E})\}$, $\hat{p}^-(E) = \max\{p^-(E), 1-p^+(\bar{E})\}$, etc.. In general, finding the improved assessment can be approached as a problem in parametric programming on the columns of the constraint matrix of the modified forms of LP1 or LP2 described above. For example, to determine $\hat{p}_i^+$, let the term $\alpha_{ij}^+ = (p_i^+ - E_{ij})F_{ij}$ be replaced by $(p_i^+ - \lambda - E_{ij})F_{ij}$ for $j=1, \ldots, m$, in the column of the constraint matrix corresponding to the variable z_i^+. The resulting linear program can be studied parametrically as a function of λ, and the improved value for p_i^+ is obtained as $\hat{p}_i^+ = p_i^+ - \lambda^*$, where λ^* is the largest value of λ for which the optimal objective value is not greater than zero.

5. General scoring rules and their probability transforms

As an alternative to betting systems, subjective probability can be defined and measured in terms of marginal rates of substitution, through the use of scoring rules. A scoring rule can be represented as a loss function of two arguments, $f(x,E)$, where E may be either 0 or 1, and x is a real number whose domain is usually taken to be the unit interval, with $f(x,0)$ being strictly increasing and $f(x,1)$ strictly decreasing in x. (That is, $x_1<x_2$ implies $f(x_1,0)<f(x_2,0)$ and $f(x_1,1)>f(x_2,1)$.) A person's subjective probability for E can be defined in terms of the value for x which he would choose under the condition of receiving a loss of $f(x,E)$. He will presumably adjust his choice for x until he finds a point at which the value for him of the marginal decrease in his loss (score) under one outcome due to further changes in x is exactly balanced by the value of the marginal increase under the other outcome. This approach can be generalized for eliciting conditional probabilities in a manner analogous to "called off bets," by letting the loss for E conditional on F be given by $f(x,E)F$, so that the loss is zero if $F=0$ obtains, regardless of the value of E.

A scoring rule is called *proper* if its x-domain is the unit interval and it has the property that a person minimizes his expected loss by choosing $x=p$ when his "true" subjective probability for E is p. The prototype proper scoring rule is the quadratic rule, $f(x,E)=k(E-x)^2$, for some constant k. This scoring rule may be considered simply as squared-error loss for choosing x to "predict" the value of E. The quadratic rule has been used by de Finetti as the basis for much of his theory of subjective probability, and also (in more general forms) has a long history of practical application in meteorology as a method of evaluating forecasts (e.g., Brier (1950), Stael von Holstein and Murphy (1978)). Another well-known proper scoring rule is the symmetric logarithmic rule, $f(x,1) = f(1-x,0) = -k(\log(x))$, whose use was recommended by Good (1952) for the reason that, with the inclusion of an appropriate additive constant, the expected reward (negative loss) to the probability assessor is proportional to the amount of information (according to Shannon's negative-entropy definition) contained in his assessment. The same effect can be obtained with respect to an assignment of probabilities to a general partition by using the asymmetric logarithmic rule: $f(x,0)=kx$, $f(x,1)=-k(\log(x))$, which is also a proper scoring rule. A detailed discussion of the properties and uses of proper scoring rules has been given by Savage (1971).

As a basis for defining and measuring subjective probability, scoring rules have an advantage over betting systems in that no intelligent antagonist is involved-- a person's net loss is determined only by the value he chooses for his probability and by the state of nature which obtains. When probabilities are elicited simultaneously for a finite number of different events on the condition that the total score will be the sum of the separate scores, the requirement of *admissibility* --that unnecessary certain loss must be avoided-- can be used to establish the same probability laws (derived from the existence of an underlying probability measure) that were established for bet prices based on the coherence requirement. De Finetti (1972, 1974) proves this result for the quadratic scoring rule by a series of geometric arguments in which the score plays the role of squared Euclidean distance. Recently Lindley (1980) has explored the properties of scales of "subjective conditional uncertainty," operationally defined in terms of generalized scoring rules satisfying only certain modest regularity requirements and whose x—domains are allowed to be arbitrary intervals of the real line. In a series of arguments based on determinants of the matrix of scoring function derivatives, somewhat parallel to de Finetti's proof of the coherence theorem, Lindley shows that admissible sets of uncertainty values elicited under a generalized scoring rule can be transformed into numbers in the unit interval which must obey the laws of probability for their respective events. In this section and the next, using appropriate definitions of admissibility and strict admissibility, a stronger version of Lindley's results will be proved by exploiting the equivalence of bet prices and choices under scoring rules, then invoking the results of the previous sections.

Let a scale of subjective conditional uncertainty be operationally defined on an interval $[x^F, x^T]$, which might be finite, semi-infinite, or infinite in extent, using a generalized scoring rule $f(x,E)$ satisfying the following regularity assumptions:

(A1) $f(\cdot,0)$ and $f(\cdot,1)$ are continuous and bounded below, and are strictly increasing and decreasing, respectively, on the interval $[x^F, x^T]$;

(A2) $f(\cdot,0)$ and $f(\cdot,1)$ have continuous first derivatives on $[x^F, x^T]$, denoted $f'(\cdot,0)$ and $f'(\cdot,1)$, respectively;

(A3) $f'(\cdot,0)$ and $f'(\cdot,1)$ are not both zero or both infinite in magnitude at any point in the open interval (x^F, x^T);

(A4) $\lim_{x \to x^F} \frac{f'(x,0)}{f'(x,1)} = 0$, and

$$\lim_{x \to x^T} \frac{f'(x,1)}{f'(x,0)} = 0.$$

These assumptions are similar, but not quite identical, to those given by Lindley (1980). Lindley's motivation for considering this generalization of the scoring-rule concept was to determine whether any method for describing uncertainty about an event or hypothesis which did not implicitly obey the laws of probability (e.g., confidence levels, fuzzy logic) could be given a subjectivistic basis in terms of a pair of loss functions. (The subsequent admissibility analysis shows this to be impossible--i.e., probability is the only sensible description of subjective uncertainty. In retrospect, this is not surprising, since choosing uncertainty values under a scoring rule is a special kind of "S-game against nature," for which there is a well-known relation between admissible strategies and Bayes strategies (Blackwell and Girshick (1954)).) One desirable property for a general uncertainty scale is monotonicity--that is, the right and left endpoints of the scale should correspond to logical (i.e, certain) truth and falsehood, respectively (hence the superscripts "T" and "F"), and for intermediate values the indicated degree of certainty (as to the event being true) should increase monotonically from left to right. To implement this notion, it appears reasonable to assume that the assigned loss should be an increasing function of x if the event turns out to be false, and a decreasing function of x if it turns out to be true. A1 is a formalization of this assumption. Another desirable property is smoothness, which is provided by the differentiability assumption, A2. A3 guarantees regularity in the sense that the losses associated with neighboring points must be distinct by a first-order amount under at least one outcome, ruling out certain kinds of degeneracy. A4 ensures that the scoring rule is well-behaved, in a sense which will be made clear below, at the endpoints of $[x^F, x^T]$. (A relaxation of A2 and A4 will be mentioned at the end of the next section.)

After a person reveals his subjective conditional uncertainty about E given F by the number x he chooses subject to a loss of $f(x,E)F$, a unique number in the unit interval, suggestive of a conditional probability, can immediately be associated with x by the marginal-rate-of-substitution argument sketched above. (It will be shown later that the numbers so determined must indeed obey the laws of probability if unnecessary certain loss is to be avoided.) In the vicinity of x, a change of Δx leads to a gain of approximately $-f'(x,1)\Delta x$ if $EF=1$ and a loss of approximately $f'(x,0)\Delta x$ if $(1-E)F=1$. (Note that, by A1, $f'(x,1) \leqslant 0$ and $f'(x,0) \geqslant 0$ for all x in $[x^F, x^T]$.) If it is assumed that the person is indifferent between x and $x+\Delta x$, regardless of whether Δx is positive or negative (provided it is sufficiently small), then his conditional subjective probability for E given F, denoted p, evidently satisfies $p(-f'(x,1))=(1-p)f'(x,0)$, leading to the equation $p=P(x)$, where $P(x)$ is the "probability transform of x" as defined by Lindley (1980):

$$P(x) \equiv \frac{f'(x,0)}{f'(x,0)-f'(x,1)}. \tag{5.1}$$

The person who chooses x to denote his uncertainty about E given F, under the scoring rule f, is therefore considered to be like the bookie who will accept an arbitrary small bet, z (either positive or negative--i.e., buying or selling), at price p, where $p=P(x)$ and $z=-(f'(x,0)-f'(x,1))\Delta x$.

Another way to interpret the probability transform is to note that if the person's "true" subjective probability for E given F is p, then his conditional expected score due to the choice x is given by

$$r(x,p) = pf(x,1) + (1-p)f(x,0). \tag{5.2}$$

A Bayesian would wish this quantity to be minimal, and a necessary condition for x to minimize $r(\cdot,p)$ is that $r'(x,p)=0$, which leads to $p=P(x)$ as defined above. (Note: "prime" will consistently denote differentiation with respect to the first argument.) Thus, x is a stationary point--either a local minimum, local maximum, or inflection point--of $r(\cdot,p)$, when $p=P(x)$, by the definition (5.1). The regularity conditions for f can be interpreted as ensuring that P is a continuous function which satisfies $0 \leqslant P(x) \leqslant 1$ on $[x^F,x^T]$, and furthermore, in view of (A4), $P(x^F)=0$ and $P(x^T)=1$. In fact, (A4) guarantees that $r(\cdot,p)$ is minimized at x^F *only* for $p=0$, and similarly at x^T for $p=1$. For any $p>0$, $r(\cdot,p)$ is minimized at some $x>x^F$, although for p sufficiently small this point can be made arbitrarily close to x^F, and similarly for $p<1$ in the vicinity of x^T.

If the probability transform is a strictly increasing function of x, then every x in $[x^F,x^T]$ is the unique point which minimizes $r(\cdot,p)$, where $p=P(x)$. A sufficient but not necessary condition for $P(x)$ to be strictly increasing is for $f(\cdot,E)$ to be strictly convex for both values of E. If $f(\cdot,0)$ and $f(\cdot,1)$ have continuous second derivatives on $[x^F,x^T]$, then P has a continuous first derivative, given by

$$P'(x) = \frac{f'(x,0)f''(x,1)-f'(x,1)f''(x,0)}{(f'(x,0)-f'(x,1))^2}. \tag{5.3}$$

In this case a necessary and sufficient condition for P to be strictly increasing is $P'(x) > 0$ almost everywhere on $[x^F,x^T]$. From the above expression this condition is seen to be equivalent to $f'(x,0)f''(x,1)-f'(x,1)f''(x,0)>0$ a.e. on $[x^F,x^T]$. (Note that this condition is weaker than strict convexity--i.e., weaker than requiring $f''(x,0)>0$ and $f''(x,1)>0$ a.e. on $[x^F,x^T]$.) Moreover, the conditional expected score function, $r(x,p)$, has a second derivative with respect to x in this case, and it is easily shown that $r''(x,p)=P'(x)(f'(x,0)-f'(x,1))$, where $p=P(x)$. By assumption A3, $f'(x,0)-f'(x,1)>0$ on (x^F,x^T), so that $r''(x,p)$ has the same sign as $P'(x)$. Hence, $r(\cdot,p)$ is locally convex at x if P is increasing at x, which implies that x is at least a local (if not global) minimum of $r(\cdot,p)$ for $p=P(x)$; and conversely, if P is decreasing at x, then x is a local maximum of $r(\cdot,p)$. This suggests that, if P is not strictly increasing on $[x^F,x^T]$, it would be sensible, in choosing x, to restrict attention to those values in whose vicinity P is increasing, insofar as a person who wished to minimize his expected score for any given probability would not do otherwise.

A proper scoring rule is a special case of a of scoring rule with a strictly increasing probability transform, for which $x^F=0$, $x^T=1$, and $P(x)=x$ on $[0,1]$. Conversely, every scoring rule with a strictly increasing probability transform can be converted into a proper scoring rule by an appropriate transformation of the x-axis--in particular, by the one-to-one transformation of $[x^F,x^T]$ onto $[0,1]$ defined by P. That is, if f is a scoring rule whose probability transform, P, is strictly increasing (and hence invertible) on $[x^F,x^T]$, then an associated proper scoring rule, denoted f^*, can be defined according to $f^*(x,E) = f(P^{-1}(x),E)$. Thus scoring rules with strictly increasing probability transforms appear to be a natural generalization of proper scoring rules, a notion which will be made more concrete in the next section.

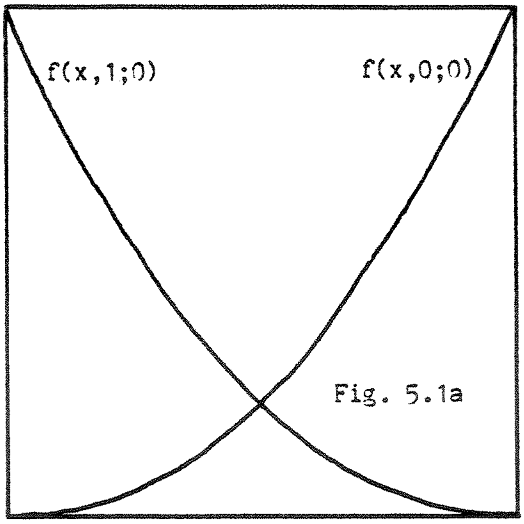
An interesting, if somewhat pathological, example of a scoring rule whose probability transform need not be strictly increasing is the trigonometric scoring rule with frequency parameter k (a non-negative integer), given by:

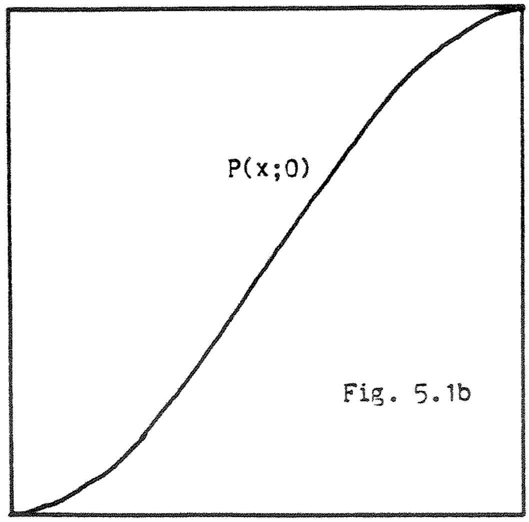
$$f(x,0;k) = f(1-x,1;k) = x - \frac{\sin(2k+1)\pi x}{(2k+1)\pi} \tag{5.4}$$

with $x^F=0$ and $x^T=1$. Note that $f(x,0;k)$ and $f(x,1;k)$ are reflections of each other in the line $x=\frac{1}{2}$, and also $f(x,0;k)=f(x,1;k)+2x-1$, whence $f'(x,0;k) - f'(x,1;k) = 2$, for all x. The corresponding probability transform is thus given by

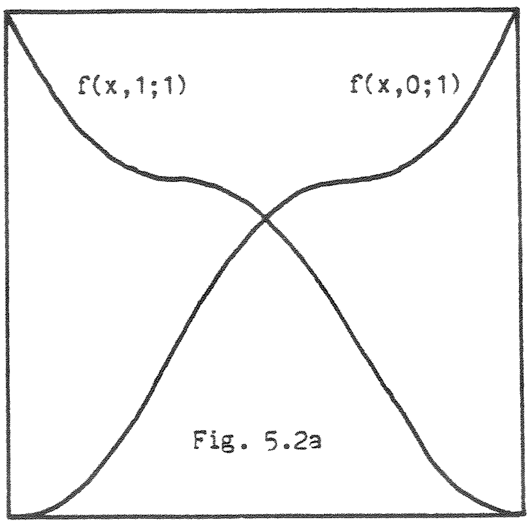
$$P(x;k) = \frac{1}{2}f'(x,0;k) = \frac{1}{2}(1-\cos(2k+1)\pi x)\,. \tag{5.5}$$

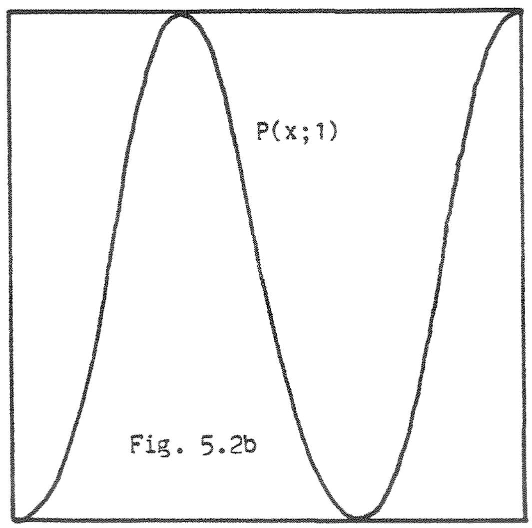
The trigonometric scoring functions, $f(x,0;k)$ and $f(x,1;k)$, and their probability transform, $P(x;k)$, are plotted in Figures (5.1a) and (5.1b) for $k=0$, and in Figures (5.2a) and (5.2b) for $k=1$. Note that the graph of the probability transform has the shape of a raised, inverted cosine wave which executes $k+\frac{1}{2}$ cycles in the unit interval. For $k=0$ the probability transform is strictly increasing and "nearly proper," i.e., its graph is close to the line $y=x$. In fact, the scoring functions closely resemble those of the quadratic rule in this case. For $k=1$, however, the probability transform increases monotonically from 0 to 1 on the interval $[0,\frac{1}{3}]$, then decreases monotonically from 1 back to 0 on $[\frac{1}{3},\frac{2}{3}]$, and finally increases monotonically from 0 to 1 again on $[\frac{2}{3},1]$. Thus, the equation $P(x;1)=p$ has two distinct solutions in x for $p=0$ or $p=1$, and three distinct solutions in x for all intermediate values of p. In particular, if x lies in the interval $[0,\frac{1}{3}]$ and is a solution to $P(x;1)=p$ for some p, then $\frac{2}{3}-x$ and $\frac{2}{3}+x$ are also solutions. Moreover, if $0<p<1$ and $p\neq\frac{1}{2}$, then the solution to $P(x;1)=p$ which lies either in $(0,\frac{1}{6})$ or in $(\frac{5}{6},1)$ is the unique global minimum of the conditional expected score function; the solution which lies either in $(\frac{1}{6},\frac{1}{3})$ or in $(\frac{2}{3},\frac{5}{6})$ is a local but not global minimum; and the solution which lies in $(\frac{1}{3},\frac{2}{3})$ is a local maximum. (Note that this is consistent with the earlier observation that if $P(x)=p$, then x is a local minimum [maximum] of $r(\cdot,p)$ if P is increasing [decreasing] at x.) The equation $P(x;1)=\frac{1}{2}$ has the solutions $x=\frac{1}{6}$, $x=\frac{5}{6}$, and $x=\frac{1}{2}$, the first two of which are both global minima of $r(\cdot,\frac{1}{2};1)$ and the last of which is the global maximum. $r(\cdot,p;1)$ is plotted in Figures (5.3a) and (5.3b) for $p=\frac{1}{4}$ and $p=\frac{1}{2}$, respectively.



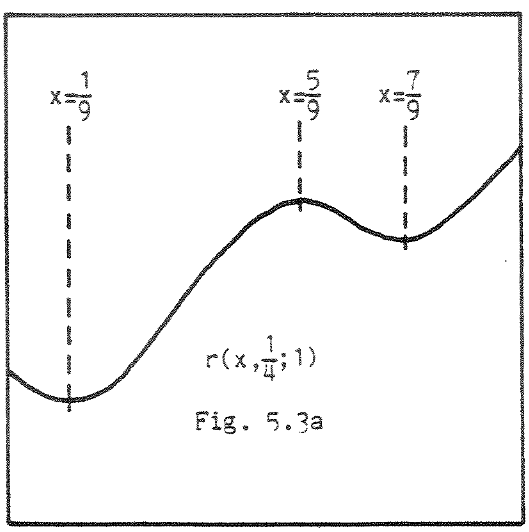


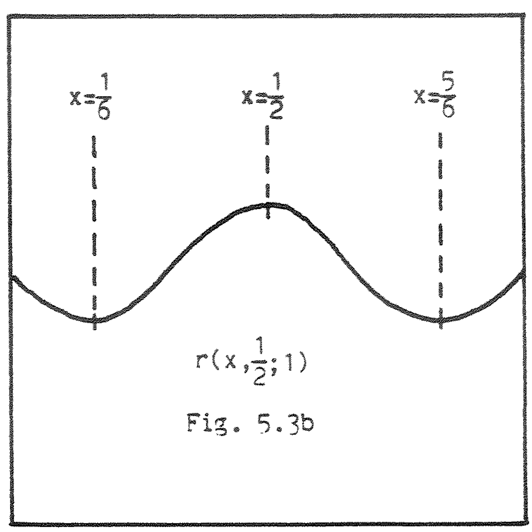
Trigonometric Scoring Rule & Probability Transform for k=0 (above) & k=1 (below)





Trigonometric Expected Score Function (below) for $p=\frac{1}{4}$ and $p=\frac{1}{2}$





6. Admissibility under generalized scoring rules

For the same n event pairs and same sample space considered earlier, let a person reveal his subjective conditional uncertainty about E_i given F_i by choosing a number x_i in an interval $[x_i^F, x_i^T]$ under a generalized scoring rule f_i satisfying the regularity assumptions given in the last section, for $i=1, \ldots, n$. (A different scoring rule may be used for each event pair. The corresponding probability transforms and conditional expected score functions will be denoted P_i, r_i, etc..) It is assumed that the person's total score (loss) under the j^{th} outcome in the sample space, denoted $s_j(\mathbf{x})$, is given by the sum of the scores for the separate event pairs, i.e.,

$$s_j(\mathbf{x}) \equiv \sum_{i=1}^{n} f_i(x_i, E_{ij}) F_{ij} \tag{6.1}$$

where the vector $\mathbf{x}$ is used to represent the set of choices $(x_1, \ldots, x_n)$. The *score vector* can now be defined as the m-vector $\mathbf{s}(\mathbf{x})$ whose j^{th} element is $s_j(\mathbf{x})$.

DEFINITION: The vector of choices $\mathbf{x}$ is *admissible* if there does not exist any other vector $\mathbf{y}$ for which $\mathbf{s}(\mathbf{y}) - \mathbf{s}(\mathbf{x}) < 0$.

This definition follows de Finetti (1972). Admissible choices under scoring rules are analogous to coherent bet prices in avoiding unnecessary uniform loss under all outcomes in the sample space; however, the corresponding notion of *strict* admissibility is not obtained merely by substituting "$\leqslant$" for "$<$" in the above definition, for reasons which will become apparent. Instead, the following is required:

DEFINITION: The vector of choices $\mathbf{x}$ is *strictly admissible* if there exists some $\epsilon > 0$ for which

$$\max_j \, [s_j(\mathbf{y}) - s_j(\mathbf{x})] \geqslant \epsilon \max_j \, [s_j(\mathbf{x}) - s_j(\mathbf{y})]$$

for all other vectors $\mathbf{y}$.

In other words, a choice vector is admissible if there is no alternative choice yielding a lower score under every outcome, and strictly admissible if, for some ϵ, every alternative choice which lowers the score by, say, Δs under one outcome, raises the score by at least $\epsilon \Delta s$ under some other outcome. Strictly admissible choices are analogous to strictly coherent bet prices in that, relative to alternative choices, they do not admit a loss under one outcome without a *proportional* gain under some other outcome. That is, a person who adheres to strict admissibility will not accept the chance of a finite [infinite] loss in return for the chance of an infinitesimal [finite] gain.

THEOREM 3: A vector of choices $\mathbf{x}=(x_1, \ldots, x_n)$ is [strictly] admissible only if $\mathbf{p}=(p_1, \ldots, p_n)$ is a [strictly] coherent vector of bet prices for the same events, where $p_i=P_i(x_i)$, $i=1, \ldots, n$.

Proof: Suppose $\mathbf{p}$ is not [strictly] coherent. Then there exists a "sure-win" ["can't-lose"] bet $\mathbf{z}$-- i.e., for which $\mathbf{t}(\mathbf{z};\mathbf{p})<0$ [$\leqslant 0$], where $\mathbf{t}(\mathbf{z};\mathbf{p})$ is the payoff vector defined in Section 2. This bet vector, together with a small positive constant δ, will be used to define a vector of small changes, $\Delta\mathbf{x}(\delta)$, such that for small enough δ the existence of the alternative choice $\mathbf{y}=\mathbf{x}+\Delta\mathbf{x}(\delta)$ will contradict the assumption of [strict] admissibility. Let $\Delta\mathbf{x}(\delta)$ be defined in the following way: if $f_i'(x_i,0)$ and $f_i'(x_i,1)$ are both non-zero and finite in magnitude, let

$$\Delta x_i(\delta) = \frac{\delta z_i}{f_i'(x_i,0)-f_i'(x_i,1)}. \tag{6.2}$$

Then,

$$f_i(x_i+\Delta x_i(\delta),E) - f_i(x_i,E) = \delta(p_i-E)z_i + o(\delta), \tag{6.3}$$

for both values of E, where "little-o" notation is used to denote an arbitrary function satisfying

$$\lim_{\delta\to 0} \frac{o(\delta)}{\delta} = 0.$$

If at least one of the derivatives is zero or infinite, then either $P_i(x_i)=0$ or $P_i(x_i)=1$. Suppose that $P_i(x_i)=0$. Then, if $f_i(x_i,1)=\infty$ (which is only possible if $x_i=x_i^F$), let $\Delta x_i(\delta)$ be chosen so that $f_i(x_i+\Delta x_i(\delta),0)-f_i(x_i,0)=\delta^2$. (This is possible, for small enough δ, since $f_i(x,0)$ is continuous and strictly increasing.) Note that, in this case, for any $\Delta x_i(\delta)>0$, $f_i(x_i+\Delta x_i(\delta),1)-f_i(x_i,1) = -\infty$, so that an infinite decrease in the score is obtained when $E_i F_i = 1$. If, however, $f_i(x,1) < \infty$, then let $\Delta x_i(\delta)$ be chosen so that $f_i(x_i+\Delta x_i(\delta),1) - f_i(x_i,1) = -\delta z_i$. (Here it may be assumed, w.l.o.g., that $z_i > 0$, since the bettor thereby obtains at least a "can't-lose" situation for the i^{th} event pair when $p_i = 0$.) Note that when $E_i F_i = 1$, the decrease in score due to $\Delta x_i(\delta)$ is

$$f_i(x_i,1) - f_i(x_i+\Delta x_i(\delta),1) = \int_{x_i}^{x_i+\Delta x_i(\delta)} (-f_i'(x,1))\,dx = \delta z_i, \tag{6.4}$$

whereas the increase in score when $(1-E_i)F_i = 1$ is

$$f_i(x_i+\Delta x_i(\delta),0) - f_i(x_i,0) = \int_{x_i}^{x_i+\Delta x_i(\delta)} f_i'(x,0)\,dx, \tag{6.5}$$

Now, $P_i(x_i) = 0$ implies that

$$\lim_{\Delta x_i(\delta)\to 0} \frac{f_i'(x_i+\Delta x_i(\delta),0)}{f_i'(x_i+\Delta x_i(\delta),1)} = 0. \tag{6.6}$$

It follows that the first integrand above can be made uniformly larger than the second by an arbitrarily large multiplicative factor by taking $\Delta x_i(\delta)$ small enough, which in turn can be accomplished by taking δ small enough, since $\Delta x_i(\delta) \to 0$ as $\delta \to 0$ by the assumed continuity and finiteness of $f_i(x,1)$ near $x = x_i$ in this case. Since the first integral is by definition proportional to δ, the second integral must therefore be proportional to $o(\delta)$. Let corresponding definitions be made for $\Delta x_i(\delta)$ if $P_i(x_i) = 1$. In this manner, a vector of changes is obtained for which $\Delta x_i(\delta) > 0$ [< 0] if-and-only-if $z_i > 0$ [< 0]. Furthermore, for both values of E, either

$$f_i(x_i+\Delta x_i(\delta),E) - f_i(x_i,E) = -\infty, \tag{6.7}$$

or,

$$f_i(x_i+\Delta x_i(\delta),E) - f_i(x_i,E) = \delta(p_i - E)z_i + o(\delta). \tag{6.8}$$

Therefore, for every j, either

$$s_j(\mathbf{x}+\Delta\mathbf{x}(\delta)) - s_j(\mathbf{x}) = -\infty, \tag{6.9}$$

or,

$$s_j(\mathbf{x}+\Delta\mathbf{x}(\delta)) - s_j(\mathbf{x}) = \delta t_j(\mathbf{z};\mathbf{p}) + o(\delta). \tag{6.10}$$

Now, by the assumption that $\mathbf{z}$ is a "sure-win" ["can't-lose"] bet, $\mathbf{t}(\mathbf{z};\mathbf{p}) < 0$ [≤ 0]; so that by taking δ small enough the score change can be made negative under every outcome [negative and proportional to δ under at least one outcome, and proportional to $o(\delta)$ under the remaining outcomes] which proves that **x** is not [strictly] admissible.

Thus it is seen that a necessary condition for a set of choices to be [strictly] admissible (in fact, locally so) is for the probability transforms to be [strictly] coherent bet prices, which in turn implies the existence of a probability distribution, **w**, and a measure based on it, $\mathbf{P_w}$, for which these are conditional probabilities. Considering the score due to the choice **x** as a random variable, denoted $S(\mathbf{x})$, the expected value of $S(\mathbf{x})$ under the probability measure $\mathbf{P_w}$ is given by

$$\mathbf{E_w}(S(\mathbf{x})) = \mathbf{w}'\mathbf{s}(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{P_w}(F_i) r_i(x_i, \mathbf{P_w}(E_i | F_i)) \tag{6.11}$$

for all **x**, where $r_i(x,p)$ is the expected partial score function defined in the last section. If **x** is a vector whose probability transform is consistent with the distribution **w**--i.e., for which $P_i(x_i) = \mathbf{P_w}(E_i | F_i)$ for every i--then $r_i'(x_i, \mathbf{P_w}(E_i | F_i)) = 0$, so that all the derivatives of the expected score function, evaluated at **x**, are zero. Theorem 3 can therefore be paraphrased as follows: **x** is [strictly] admissible only if there exists a [positive] probability distribution on the sample space for which the gradient of the expected score, evaluated at **x**, is the zero vector. The gradient being the zero vector is, of course, a first-order necessary condition for an unconstrained minimum of a smooth function. Thus, a necessary condition for [strict] admissibility is that **x** must satisfy a first-order condition for a minimum of the expected score, under a [positive] probability distribution on the sample space. On the other hand, consideration of the properties of a minimum of the expected score leads to sufficient conditions for admissibility or strict admissibility. A choice **x** which minimizes the expected score under the distribution **w** is said to be "Bayes against **w**." **x** will simply be described as *Bayes* if it is Bayes against some **w**, and *strictly Bayes* if it is Bayes against some $\mathbf{w} > 0$. In these terms, we have:

THEOREM 4: **x** is [strictly] admissible if it is [strictly] Bayes.

Proof: For the non-strict case this result is obvious-- an alternative choice yielding a lower score under every outcome would also yield a lower expected score under every probability distribution on the sample space. For the strict case, assume **x** is Bayes against some positive probability distribution, and let w_{min} denote the least element of this distribution. Then suppose that **x** is not strictly admissible, i.e., that for every positive ϵ, no matter how small, there exists an alternative choice which lowers the score by, say, Δs under one outcome without raising the score by more than $\epsilon\Delta s$ under any other outcome. Thus the score can be lowered by Δs under some outcome with probability greater than or equal to w_{min}. Under all other outcomes, the score is raised (if at all) by not more than $\epsilon\Delta s$, and the total probability of these other outcomes cannot be more than $1-w_{min}$. By choosing $\epsilon < \frac{w_{min}}{(1-w_{min})}$, this alternative choice can therefore be made to have a lower expected score than **x**.

THEOREM 5: If, for every i, the probability transform P_i is strictly increasing on $[x_i^F, x_i^T]$, and $p_i = P_i(x_i)$, then the following are equivalent:

(i) **p** is [strictly] coherent;

(ii) **x** is [strictly] Bayes;

(iii) **x** is [strictly] admissible.

Proof: Suppose **p** is [strictly] coherent. Then, by Theorem 2, there exists a [positive] probability distribution, **w**, for which $p_i = \mathbf{P}_\mathbf{w}(E_i | F_i)$, or else $\mathbf{P}_\mathbf{w}(F_i) = 0$, for every i. Since $p_i = P_i(x_i)$ and P_i is strictly increasing on $[x_i^F, x_i^T]$, x_i uniquely minimizes $r_i(\cdot, p_i)$. From the representation of the expected total score given in Equation (6.11), it follows that **x** is Bayes against **w**. Thus, **x** is [strictly] Bayes if **p** is [strictly] coherent. By Theorem 4, **x** is [strictly] admissible if it is [strictly] Bayes. Finally, by Theorem 3, **p** is [strictly] coherent if **x** is [strictly] admissible.

This theorem completes the generalization of de Finetti's notion of the equivalence of betting systems and scoring rules as methods for defining and measuring subjective probability, which was recently proved by Lindley (1980) in a weaker form. It has been shown that the requirements of [strict] admissibility for conditional uncertainty assessments under generalized scoring rules with strictly increasing probability transforms give rise to the probability laws in the same way as the requirements of [strict] coherence for conditional bet prices. If, however, the probability transforms are not all strictly increasing, then the existence of a probability measure consistent with the probability transforms of the choices, which is a necessary condition for admissibility according to Theorem 3, is not also a sufficient condition. An example of the latter situation is provided by the trigonometric scoring rule introduced previously, for the case in which $k=1$. For some event E, let x_1 and x_2 be chosen to describe the unconditional uncertainty of E and $1-E$, respectively, both under this scoring rule. For either value of E, the total score is then given by $f(x_1,E;1)+f(x_2,1-E;1)$, where $f(x,E;k)$ is defined by Equation (5.4). From Theorem 3, a necessary condition for (x_1,x_2) to be admissible is $P(x_1;1)+P(x_2;1)=1$, where $P(x;k)$ is given by Equation (5.5). For some p in $(0,1)$, consider all pairs (x_1,x_2) which meet the above condition by satisfying $P(x_1;1)=p$ and $P(x_2;1)=1-p$. There are nine such distinct pairs, corresponding to the combinations of the three solutions for x_1 and the three solutions for x_2, as noted at the end of Section 5. If $p\neq\frac{1}{2}$, exactly one of these nine pairs is Bayes against $\mathbf{w}=(p,1-p)$, namely the unique pair of which one element lies in $(0,\frac{1}{6})$ and the other element lies in $(\frac{5}{6},1)$. (This pair is admissible, by Theorem 4.) Also, exactly one pair is inadmissible, namely the unique pair of which both elements lie in $(\frac{1}{3},\frac{2}{3})$. (This can be demonstrated by a simple geometrical argument, based on the fact that both $f(x,0;1)$ and $f(x,1;1)$ are concave on $(\frac{1}{3},\frac{2}{3})$.) The remaining seven pairs are also, in fact, admissible, even though none of them is Bayes against any $\mathbf{w}$. If $p=\frac{1}{2}$ there are four admissible pairs which are Bayes $(x_1,x_2 \in \{\frac{1}{6}, \frac{5}{6}\})$, one inadmissible pair $(x_1=x_2=\frac{1}{2})$, and four admissible pairs which are not Bayes.

The above results can be extended to the case in which assumptions A2 and A4 are relaxed to allow scoring functions whose derivatives are only piecewise continuous. In particular, assume that $f_i(\cdot,0)$ and $f_i(\cdot,1)$ have piecewise continuous derivatives such that the corresponding probability transform is piecewise continuous, and satisfies $P_i^-(x) \leq P_i^+(x)$ for all x in $[x_i^F, x_i^T]$, and also $P_i^+(x_i^F) < 1$ and $P_i^-(x_i^T) > 0$, where $P_i^-(x)$ and $P_i^+(x)$ denote the limits from the left and right, respectively, of P_i at x. Then, by the same same procedure as in the proof of Theorem 3, invoking the results of Theorem 2' rather than Theorem 2, it can be shown that a choice $\mathbf{x}$ is [strictly] admissible only if $(\mathbf{p}^-, \mathbf{p}^+)$ is a set of [strictly] coherent buying/selling bet prices, where $p_i^- = P_i^-(x_i)$ and $p_i^+ = P_i^+(x_i)$ for all i, with $P_i^-(x_i^F) \equiv 0$ and $P_i^+(x_i^T) \equiv 1$. If P_i is also strictly increasing for every i, then a corresponding generalization of Theorem 5 is obtained. Thus, upper and lower probabilities can also arise in the context of scoring rules.

Blackwell and Girshick (1954) give numerous admissibility results for statistical games, using a definition of admissibility which is intermediate in strength between admissibility and strict admissibility as defined in this paper. (In particular, their definition of admissibility, which follows Wald (1950), is obtained by substituting "$\geqslant$" for "$>$" in the definition of admissibility at the beginning of this section.) An "S-game against nature" is a statistical game defined by a set S in m-space, in which the player chooses a strategy consisting of a vector $\mathbf{s} = (s_1, \ldots, s_m)$ in S, and "nature" then randomly chooses a coordinate, j, whereupon the player receives a loss of s_j. The process of eliciting conditional uncertainty assessments for a set of event pairs under generalized scoring rules, as described in this section, is clearly a special kind of S-game against nature, in which the set S consists of all $\mathbf{s}(\mathbf{x})$ generated by (6.1) for values of $\mathbf{x}$ satisfying $x_i^F \leqslant x_i \leqslant x_i^T$, $i=1, \ldots, n$. This set is closed, but not generally convex; although if the probability transforms of the scoring rules are all strictly increasing, then the admissible points lie on a convex boundary. Blackwell and Girshick show that for S-games in which S is closed and convex, with the "$\leqslant$" definition of admissibility, every strictly Bayes strategy is admissible, and every admissible strategy is Bayes. This result is applicable to the case of scoring rules with strictly increasing probability transforms, but it is less specific than Theorem 5. However, by a direct application of the basic separation theorem, it can be shown that in every S-game which is closed, convex, and bounded below, a strategy is [strictly] admissible if-and-only-if it is [strictly] Bayes, under the definitions used here.

7. References

Blackwell, David, and Girshick, M.A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

Brier, G.W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review 78* 1-3

Buehler, R.J. (1976). Coherent Preferences. *Ann. Statist. 4* 1051-1064

Cornfield, J. (1969). The Bayesian outlook and its applications. *Biometrics 25* 617-658

Dawid, A.P. and Stone, M. (1972). Expectation consistency of inverse probability distributions. *Biometrika 59* 486-489

Dawid, A.P. and Stone, M. (1973). Expectation consistency and generalized Bayes inference. *Ann. Statist. 1* 478-485

de Finetti, Bruno (1937). Foresight: Its logical laws, its subjective sources. Reprinted in Kyburg, H., and Smokler, H., eds. (1964). *Studies in Subjective Probability.* Wiley, New York. 93-158

de Finetti, Bruno (1972). *Probability, Induction, and Statistics.* Wiley, New York.

de Finetti, Bruno (1974). *Theory of Probability, 1.* Wiley, New York.

Dempster, A.P. (1968). A generalization of Bayesian inference. *J. Roy. Stat. Soc. B, 40* 205-247

Dunford, Nelson, and Schwartz, Jacob T. (1958). *Linear Operators, Part I: General Theory.* Interscience, New York.

Freedman, David A. and Purves, Roger A. (1969). Bayes' method for bookies. *Ann. Math. Stat. 40* 1419-1429

Good, I.J. (1952). Rational decisions. *J. Roy. Stat. Soc. B, 14,* 107-114

Heath, David C. and Sudderth, William D. (1972). On a theorem of de Finetti, oddsmaking, and game theory. *Ann. Math. Stat. 43* 2072-2077

Heath, David C. and Sudderth, William D. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Stat. 6* 333-345

Koopman, B.O. (1940). Axioms and algebra of intuitive probability. *Annals of Mathematics 41* 269-292

Lindley, D.V. (1980). Scoring rules and the inevitability of probability. *ORC* 81-1. Operations Research Center, University of California, Berkeley.

Lindley, D.V., Tversky, A. and Brown, R.V. (1979). On the reconciliation of probability assessments. *J. Roy. Stat. Soc. A, 142* Part 2, 146-180

Pierce, Donald A. (1973). On some difficulties in a frequency theory of inference. *Ann. Stat. 1* 241-250

Savage, L.J. (1971). Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc. 66* 783-801

Smith, C.A.B. (1961). Consistency in statistical inference and decision. *J. Roy. Stat. Soc. B, 23* 1-25

Stael von Holstein, Carl-Axel S. and Murphy, Allan H. (1978). The family of quadratic scoring rules. *Monthly Weather Review 106* 917-924

Wald, A. (1950). *Statistical Decision Functions.* Wiley, New York.