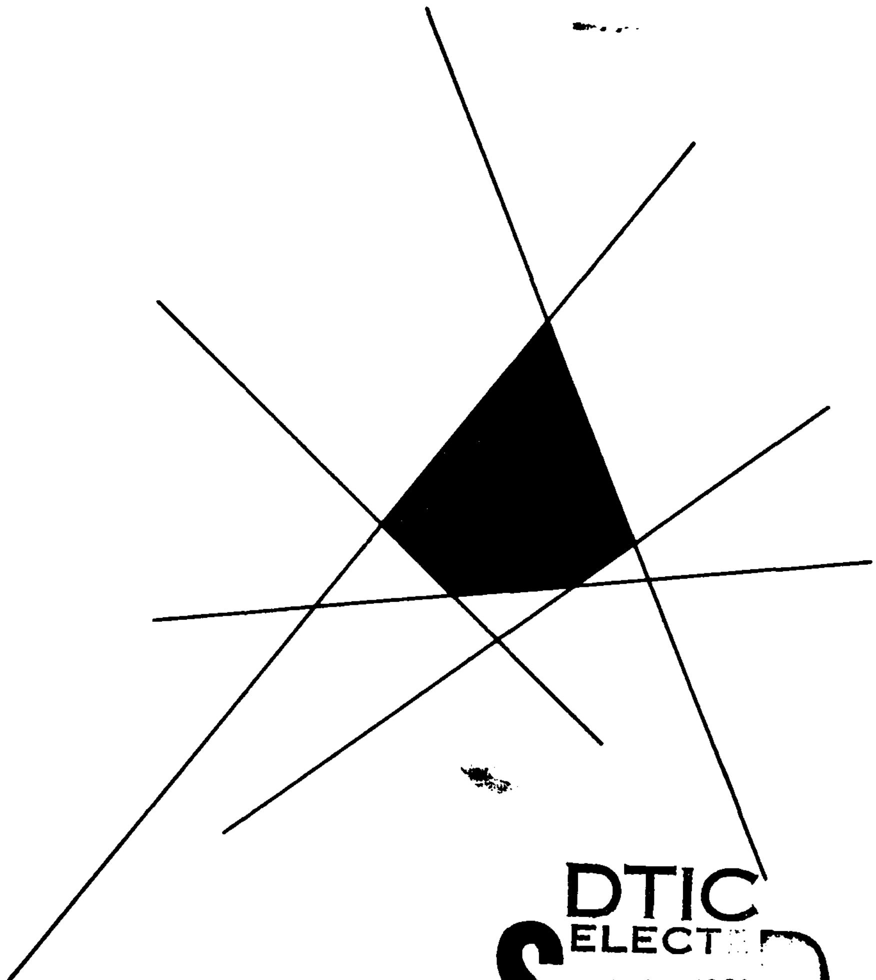
ORC 81-1
JANUARY 1981

LEVEL II

SCORING RULES AND THE INEVITABILITY OF PROBABILITY

by

DENNIS V. LINDLEY



OPERATIONS
RESEARCH
CENTER

UNIVERSITY OF CALIFORNIA · BERKELEY

SCORING RULES AND THE INEVITABILITY OF PROBABILITY[†]

by

Dennis V. Lindley[††]

JANUARY 1981 ORC 81-1

[†]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. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[††]Present address: 2 Periton Lane, Minehead, Somerset TA24 8AQ, England.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC 81-1	2. GOVT ACCESSION NO. AD-A097045	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SCORING RULES AND THE INEVITABILITY OF PROBABILITY		5. TYPE OF REPORT & PERIOD COVERED Research Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dennis V. Lindley		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 Jan[redacted]81
		13. NUMBER OF PAGES 26
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)

Scoring Rules; Confidence Statements; Significance Tests; Fuzzy Logic; Possibilities; Upper and Lower Probabilities; Admissibility; Proper Rules; Finite Additivity; Expectation; Pareto Optimality.

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 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

Let a person express his uncertainty about an event E, conditional upon an event F, by a number x and let him be given, as a result, a score which depends on x and the truth or falsity of E when F is true. It is shown that if the scores are additive for different events and if the person chooses admissible values only, then there exists a known transform of the values x to values which are probabilities. In particular, it follows that values x derived by significance tests, confidence intervals or by the rules of fuzzy logic are inadmissible. Only probability is a sensible description of uncertainty.

SCORING RULES AND THE INEVITABILITY OF PROBABILITY

by

Dennis V. Lindley

Introduction

Suppose that a person, considering an event E about which he is uncertain, describes that uncertainty by a number x. De Finetti's (1974) basic argument is that if the person is scored an amount $(x - 1)^2$ if E is true and x^2 if E is false, and if the scores for different events are additive, then x must be a probability for E. This result has been generalized to some other scores besides the quadratic one: a seminal paper is that by Savage (1971) which contains several references. In the present paper we show that De Finetti's argument applies to virtually every reasonable score function with the only modification that a known transform of x, rather than x itself, must be a probability. The argument may be viewed as providing another axiomatic justification for the Bayesian position, advantages being the simplicity both of the assumptions and of the proof. It also demonstrates that any description of uncertainty by numbers that do not obey the rules of the probability calculus, even after transformation, will violate the simple assumptions we make. Examples of such non-probabilistic assignments are significance levels, confidence statements and possibilities in fuzzy logic. The argument is extended to where the description is by means of two numbers, perhaps upper and lower probabilities, as suggested by Dempster (1968) and Smith (1961), to demonstrate that these are in disagreement with the assumptions. The message is essentially that only probabilistic descriptions of uncertainty are reasonable.

I am grateful to Richard E. Barlow for inviting me to Berkeley and to L. A. Zadeh who asked me to give a seminar on the relationship between probability and the ideas of fuzzy logic. This seminar suggested the possibility of the existence of a scoring rule that led to the laws of fuzzy sets: the paper shows no such rule exists. The observations of Robert Nau on a first draft of the paper have been of considerable value to me.

Notation:

We consider real variables $X, Y, \ldots$ taking values $x, y, \ldots$. Events are denoted by $E, F, \ldots$ and the same symbol is used for the indicator variable of an event, so that $E = 1$ (0) if E is true (false). $f(X,E)$ is a function of the variables X and E : $f(x,1)$ is the value of that function when X takes the value x and E is true. $f'(X,E)$ is the derivative of that function with respect to X .

Score Assumption:

For a given score function $f(X,E)$, a person who describes his uncertainty about E , conditional on F , by a real number x will receive a score $f(x,E)F$. The scores are additive in that if x_i refers to E_i conditional on F_i for $i = 1,2, \ldots, n$, then the total score for all these descriptions will be $\sum_{i=1}^{n} f(x_i, E_i)F_i$.

We consider the question of what are reasonable values for him to choose. A score may be thought of as a reward or as a penalty. For definiteness we shall think of it as a penalty, so that the person wishes to reduce his score.

Admissibility Assumption:

A person will not choose values x_i for E_i conditional on F_i $(i = 1,2, \ldots, n)$ if there exist values $y_1, y_2, \ldots, y_n$ such that

$$\Sigma_i f(y_i, E_i) F_i \leq \Sigma_i f(x_i, E_i) F_i \tag{1}$$

for all values of the indicator variables, and strict inequality holds for some values.

If the conditions do obtain, he could reduce his penalty in some circumstances without increasing it in any. In statistical language the set $(x_1, x_2, \ldots, x_n)$ is inadmissible and the assumption says that only admissible values will be selected.

Origin and Scale Assumption:

For the uncertainty of E $(\bar{E})$ conditional on E, there exists a unique admissible value x_T (x_F) the same for all E; and $x_F \neq x_T$.

The suffix T (F) denotes true (false). Without loss of generality we suppose $x_F < x_T$.

Regularity Assumptions:

X can assume all values in a closed interval I of the real line. $f'(X,E)$ exists, is continuous in X for each E and, for both $E = 1$ and $E = 0$, vanishes at most once. Also x_F and x_T are interior points of I.

These regularity assumptions are unnecessarily restrictive and are later relaxed. Our reason for introducing them in this form is that the proof is then unencumbered with side-issues that might otherwise obscure the argument. We first prove three lemmas.

<u>Lemma 1</u>:

All values in the closed interval $[x_F, x_T]$ are admissible, and values outside are inadmissible. The function

$$P(x) = \frac{f'(x,0)}{f'(x,0) - f'(x,1)} \qquad (2)$$

satisfies $0 \leq P(x) \leq 1$ in $[x_F, x_T]$, is continuous and $P(x_F) = 0$, $P(x_T) = 1$. In particular the equation in x, $P(x) = p$ has at least one admissible solution for any p with $0 \leq p \leq 1$.

For E conditional on E the only score is $f(x,1)$. By the origin and scale assumption, x_T is the unique admissible value and therefore minimizes this function. By the regularity assumption $f'(x_T,1) = 0$. Similarly for $\bar{E}$ conditional on E, $f'(x_F,0) = 0$. Again by the regularity assumption $f'(x,1) > (<) \; 0$ for $x > (<) \; x_T$ and $f'(x,0) > (<) \; 0$ for $x > (<) \; x_F$: in particular, $f'(x_T,0) > 0$ and $f'(x_F,1) < 0$.

For E conditional on F the score will be $f(x,1)$ if $EF = 1$ and $f(x,0)$ if $(1 - E)F = 1$, and otherwise zero. If $f'(x,1)$ and $f'(x,0)$ are both strictly positive (negative) x is inadmissible since a small decrease (increase) in x will reduce both scores. Combining this with the result in the final sentence of the last paragraph, we see that only values in $[x_F, x_T]$ are admissible. All values in $[x_F, x_T]$ are admissible since any decrease from x, although it will lower $f(x,0)$, will necessarily increase $f(x,1)$: and similarly for an increase from x.

The properties claimed for $P(x)$ all easily follow from the results already established.

<u>Lemma 2</u>:

The values x for E and y for $\bar{E}$, both conditional on F, being admissible imply $P(x) + P(y) = 1$.

The total scores in the two possible cases will be:

$$\left.\begin{array}{ll} EF = 1 & f(x,1) + f(y,0) \\ (1 - E)F = 1 & f(x,0) + f(y,1) \,. \end{array}\right\}$$

Consider small changes in x to $(x + h)$ and y to $(y + k)$. The resulting changes in these scores will be, to order h and k,

$$\left.\begin{array}{l} f'(x,1)h + f'(y,0)k \\ f'(x,0)h + f'(y,1)k \,. \end{array}\right\}$$

and

Both these changes could be made negative, so reducing both scores and making (x,y) inadmissible, by solving the linear equations in h and k obtained by equating these to small, selected, negative values. The only exception to this occurs when the determinant of the linear equations vanishes. The condition for this is that $f'(x,1)f'(y,1) = f'(x,0)f'(y,0)$ or $P(x) + P(y) = 1$.

This argument fails at boundary points because the values of h or k required to reduce both scores may not be permissible. Consider the case $x = x_T$ where $h < 0$ and $f'(x,1) = 0$. If $y \neq x_F$, so that $f'(y,0) > 0$, the first change is $f'(y,0)k$ and for this to be negative, $k < 0$. Since $f'(x,0) > 0$, the second change can be made negative. Hence $x = x_T$ is inadmissible unless $y = x_F$ when the first change is necessarily positive and $P(x_T) + P(x_F) = 1$. Other boundary values follow similarly.

Lemma 3:

The values, x for F conditional on G, y for E conditional on FG, and z for EF conditional on G, being admissible implies $P(z) = P(x)P(y)$.

The method of proof follows that of Lemma 1. The total scores in the three possible cases will be:

$$\left.\begin{array}{rl} EFG = 1 & f(x,1) + f(y,1) + f(z,1) \\ (1 - E)FG = 1 & f(x,1) + f(y,0) + f(z,0) \\ (1 - F)G = 1 & f(x,0) \qquad\qquad\quad + f(z,0)\,. \end{array}\right\}$$

Consider small changes in x, y and z; then these can result in changes in the three total scores that are all negative, so making (x,y,z) inadmissible, unless the determinant of the linear equations is zero. Simple calculation establishes that the determinant is

$$[f'(x,0) - f'(x,1)][f'(y,0) - f'(y,1)][f'(z,0) - f'(z,1)][P(x)P(y) - P(z)]\,.$$

The first three factors do not vanish by results established in the proof of Lemma 1. Hence the last factor vanishes, as required. The boundary values require special consideration as in Lemma 2; details are omitted.

Theorem:

The four assumptions listed above imply that the values x describing uncertainty will be such that the transforms $P(x)$ obey the laws of probability.

Lemma 1 establishes the convexity property that $0 \leq P(x) \leq 1$ and $P(x_F) = 0$. Lemma 2 is the additive property. Lemma 3 is the multiplicative property.

The theorem states that admissibility implies probability, through a transform of the stated value, but not the converse. To consider this, suppose that a person has probability p for E and considers that the relevant quantity is his expected score

$$pf(x,1) + (1 - p)f(x,0) .$$

He will minimize this over x with the result that $p = P(x)$, in accord with the theorem. The same argument applies in the circumstances of Lemmas 2 and 3. Minimization of expectation gives an admissible result, so that we can state the

Corollary:

If the equation in x, $P(x) = p$, has a unique root for all $0 \le p \le 1$, then all scores such that $P(x)$ obeys the rules of probability are admissible.

If $P(x) = p$ has a unique root, we shall refer to the scoring rule as single-valued. If it has multiple roots we have the possibilities of probability rules giving inadmissible values, or of admissible values not obtained through minimization of an expectation. (Examples below show that both possibilities can occur.) Our next result enhances the status of the probability transform of x.

Lemma 4:

If, in considering the uncertainty of E conditional on F with score function $f(X,E)F$, a person gives x; and with score function $g(Y,E)F$, gives y: then $P(x) = Q(y)$.

Here $Q(y) = g'(y,0)/\{g'(y,0) - g'(y,1)\}$ see (2), and the result says that if the score function is changed the probability transform is invariant. The proof uses the method of Lemmas 2 and 3. The first-order changes will be

$$\left.\begin{array}{ll} EF = 1 & f'(x,1)h + g'(y,1)k \\ (1 - E)F = 1 & f'(x,0)h + g'(y,0)k \end{array}\right\}$$

and the determinant necessarily being zero gives $P(x) = Q(y)$.

It follows that a person could proceed by choosing his probability p in advance of knowing what score function was to be used and then, when it was announced, providing x satisfying $P(x) = p$. Robert Nau has pointed out to me that in the proof of the theorem there is no need for the score function to be the same for each event considered: each value can be transformed by its own probability transform to give a probability value. The next result shows that any probability transform is possible.

<u>Lemma 5</u>:

For any function $P(x)$ having the properties described in Lemma 1, there exists a score function with $P(x)$ as probability transform.

For example, let $f(x,0) = (x - x_F)^2$, the quadratic function. Then from (2)

$$f'(x,1) = 2(x - x_F)[P(x) - 1]/P(x)$$

which, with the boundary condition $f'(x_T,1) = 0$, yields a solution for $f(x,1)$ satisfying the regularity conditions.

In all the discussion so far the only ordering of scores has been based on admissibility and shows that the probability-transforms include all admissible values. But when a person selects a value x to describe his uncertainty he is using more than admissibility: he is selecting one value out of all admissible ones. In particular, in the case of multiple roots, he is selecting amongst x-values that yield the same probability. Our next assumptions concerns this additional ordering.

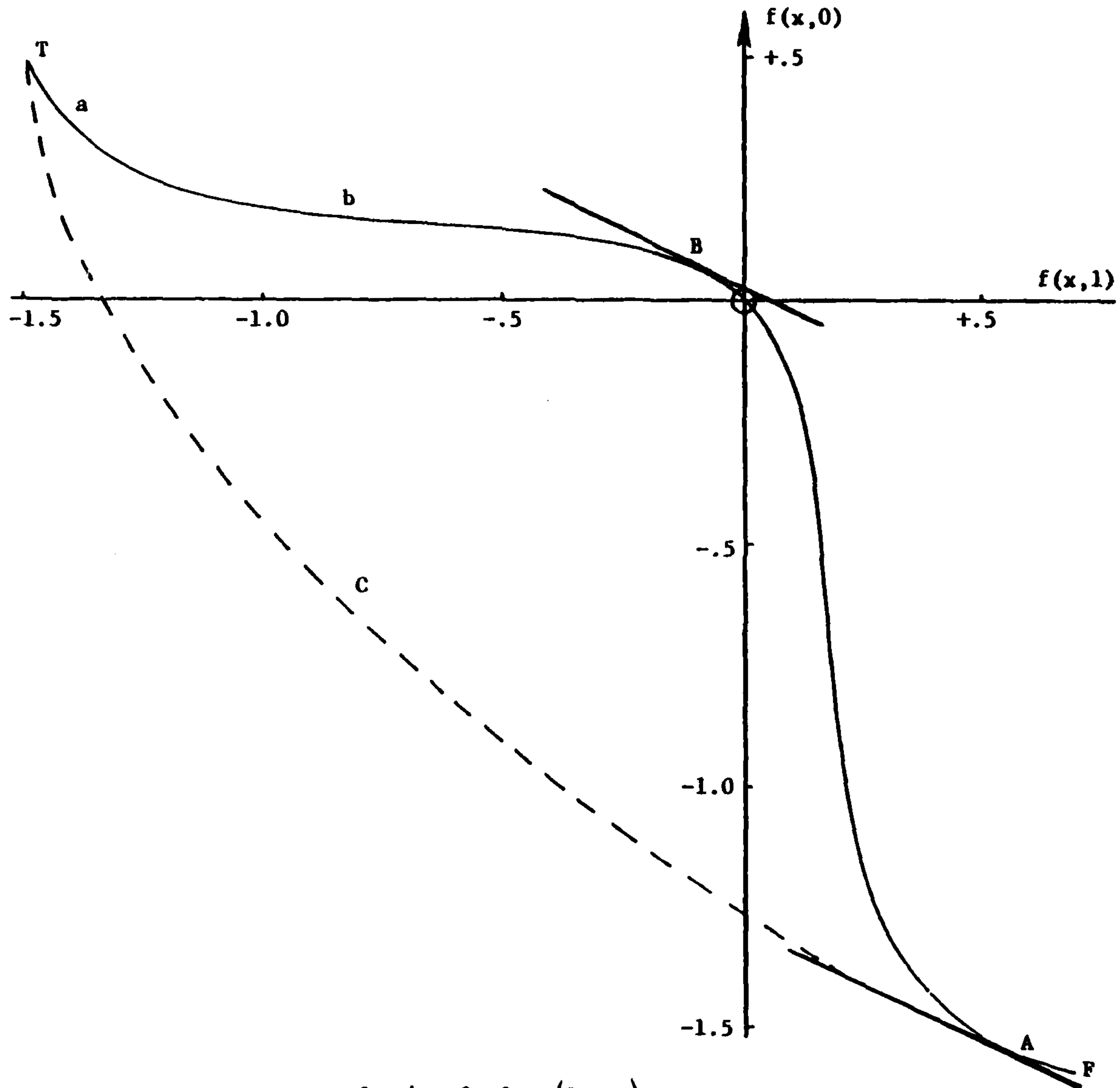
Invariance Assumption:

Any preferences amongst scores do not depend on the score function being used, and such preferences are transitive.

Theorem 2:

The five assumptions listed above imply that the values x describing uncertainty will be such that the transforms $P(x)$ obey the laws of probability and conversely that any x may be attained by selecting probabilities and minimizing the expected score.

Only the second part requires proof (the first is Theorem 1) and we use the figure. Here the axes are the scores $f(x,1)$ and $f(x,0)$ and the solid curve describes these coordinates as x varies from x_F (at F) to x_T (at T). (It is actually the curve of a quartic rule to be described below, but will serve for the proof.) This curve will have slope $f'(x,0)/f'(x,1) = -P(x)/\{1 - P(x)\}$, always negative and varying continuously from zero at F to $-\infty$ at T. By the remark above, any curve will these properties can be obtained from suitable f's. The points A and B are both admissible and have the same transforms $P(x)$--the tangents at A and B have the same slope. The dotted curve corresponds to another score function which is single-valued and passes through A with the same slope. On this curve there clearly exists a point C with both scores less than those of B.



$$f(x,E) = \frac{1}{16}x^4 - \frac{3}{8}x^2 + \left(\frac{1}{2} - E\right)x \; : \; -2 \le x \le 2$$

FIGURE 1

Now consider an event with probability p. With the single-valued score function, A is preferred to C. By admissibility C is preferred to B. Hence, by the invariance assumption, A is preferred to B with the original score function. This argument is available for any point, like A, that minimizes the expected score, and the theorem is established.

The results obtained apply only to score functions which obey both the origin and scale, and regularity assumptions. The former is not essential. One can have score functions with several minima and, in particular, several possible descriptions of a sure event. This leads to ambiguities which can be resolved in the sense that they all lead to the same probability, namely one, after transformation. No advantage seems to accrue from such flexibility.

The regularity assumptions require considerable discussion. The existence and continuity of the derivatives is introduced in order to avoid abrupt changes in the score. The nonvanishing of the derivatives, except at x_F and x_T, is a slight strengthening of the natural requirement that, at least for admissible values, the score function does not take the same value for two different choices x_1 and x_2; for if it did, there would be no rationale for choosing between x_1 and x_2 and again there would be ambiguity. The unnecessarily severe restriction is that x_F and x_T are interior points, introduced to ensure that the minima are obtained by the differential calculus, a condition that need not obtain on boundary points. We consider the case where x_F is a boundary point: an analogous treatment applies at x_T.

The major difference now is that we do not necessarily have $f'(x_F,0) = 0$. Suppose that we add the condition that $\lim_{x \to x_F} f'(x,0)/f'(x,1) = 0$ and require that $f'(x,0) > 0$ for $x > x_F$. The effect of the limit condition is to make $\lim_{x \to x_F} P(x) = 0$. It is then straightforward to show that the properties of $P(x)$ proved in Lemma 1 still obtain, as do the boundary features considered in Lemmas 2 and 3. The curve of admissible values used in the proof of Theorem 2 will still have zero slope at x_F and the argument used there carries over. Consequently both theorems remain true.

We therefore restate the

Regularity Assumptions:

X can assume all values in a closed interval I of the real line. $f'(X,E)$ exists and is continuous in X for each E . For $x > (<) x_F$, $f'(x,0) > (<) 0$: for $x > (<) x_T$, $f'(x,1) > (<) 0$. Then either x_F is an interior point of I or $\lim_{x \to x_F} f'(x,0)/f'(x,1) = 0$. Also either x_T is interior or $\lim_{x \to x_T} f'(x,1)/f'(x,0) = 0$.

Under these conditions Theorem 1 and 2 persist.

We now offer several miscellaneous comments on the results.

1. Throughout the discussion we have referred to uncertainty of E conditional on F because conditional assessment is the general form. If the person knows that F is true then we may speak of the uncertainty of E . It should be remembered that the full force of the phrase "conditional on F" is "were the person to be told that F is true". He is assessing the situation now and scores are only nonzero, and therefore, of concern to him, when $F = 1$. He need only consider the case $F = 1$ but does not need to know that $F = 1$.

2. A scoring rule is proper if it leads directly to a probability: that is, if $P(x) = x$ or

$$xf'(x,1) = (1 - x)f'(x,0) .$$

The quadratic rule used by De Finetti has $f(x,1) = (x - 1)^2$ and $f(x,0) = x^2$ and is clearly proper. As an example of an improper rule consider $f(x,1) = (x - 1)^4$ and $f(x,0) = x^4$, in which the fourth powers replace the squares of the proper rule. $P(x)$ is then $x^3/(3x^2 - 3x + 1)$ and $P(x) = p$ is a cubic in x with a unique root x for any $0 \le p \le 1$.

The quartic rule, suggested to me by Robert Nau,

$$f(x,E) = \frac{1}{16} x^4 - \frac{3}{8} x^2 + \left(\frac{1}{2} - E\right)x$$

provides an example for which $P(x) = p$ has multiple roots or is not single-valued. Here $x_F = -2$, $x_T = +2$, the regularity conditions are obeyed with these as interior points and

$$P(x) = (x + 2)(x - 1)^2/4$$

a cubic with three roots in x for every p, $0 \le p \le 1$. It is the scores for a single event with this rule that are graphed in the figure. As x decreases from $x_T = 2$ the scores move from T along the curve to the point a when $x = \sqrt{3}$. These points lie on a convex part of the curve and can be obtained by minimizing the expected score. As x decreases further the curve remains convex until at $x = 1$ it reaches the point b ; but these points, though admissible, cannot be obtained by a minimization of the expected score and are dominated by points near

F (x near x_F) having the same tangent slope. Between $x = 1$ and $x = 0$, when the curve reaches the origin, the curve is concave but the values are still admissible though again dominated by values near x_F. The situation repeats itself between the origin and F with $-x$ for x and T for F. Only values $3 \leq x^2 \leq 4$ are satisfactory and can be obtained by minimizing the expected score. Between -2 and $-\sqrt{3}$, $P(x)$ increases monotonically from 0 to 1/2 : between $+\sqrt{3}$ and $+\sqrt{2}$ it similarly passes from 1/2 to 1. The stated value has a discontinuity as p passes through 1/2. It is generally true that the condition for curvexity is $P'(x) > 0$: this obtains here with $1 \leq x^2 \leq 4$. The remaining values $|x| \leq 1$, give points on the concave part of the curve.

If the scores are plotted for E and $\tilde{E}$ (cf. Lemma 2) then the curve $P(x) + P(y) = 1$ again gives the three types of points just considered--minimizing an expected score, convex but not obtained by minimization, concave--but also points which are inadmissible. These latter arise when $|x| \leq 1$ and $y = -x$.

3. The regularity assumptions are all obviously reasonable except those on the limits at x_T and x_F when they are not interior points. Consider what happens when they do not hold, specifically suppose $\lim_{x \to x_F} f'(x,0)/f'(x,1) < 0$, or $\lim_{x \to x_F} P(x) = a > 0$. This implies x must be chosen so that $P(x) \geq a$ or is zero. But Lemma 3 shows that this implies $P(x) \geq a^{\frac{1}{2}}$, and so on. Hence all values of x must be such that $P(x)$ is 0 or 1: that is, the only admissible values are $x = x_F$ and $x = x_T$. Such score functions are trivial in that they always lead to asserting the truth or falsity of any event, a practice

which is encouraged in present-day teaching by the requirement that the pupil is always expected to answer from a dichotomy "yes" or "no": "right" or "wrong".

A strange scoring rule illustrating this is the square-root rule with $f(x,1) = (1-x)^{\frac{1}{2}}$ and $f(x,0) = x^{\frac{1}{2}}$ for $0 \leq x \leq 1$. Here $x_F = 0$, $x_T = 1$. Since $f^2(x,1) + f^2(x,0) = 1$ the curve of admissible values is the quarter of the unit circle in the positive quadrant centered at the origin, which is entirely concave. The only points that can be reached by minimizing an expected score are x_F and x_T. The regularity conditions are not satisfied: indeed $\lim_{x \to x_F} f'(x,0)/f'(x,1)$ is infinite and $P(x)$ decreases with x.

4. We now turn to scoring rules that are more useful. The logarithmic form,

$$f(x,1) = -\log x \quad \text{and} \quad f(x,0) = -\log(1-x) ,$$

is defined only in $[0,1]$. It is proper with $P(x) = x$. The hyperbolic form

$$f(x,1) = x^{-1} \quad \text{and} \quad f(x,0) = (1-x)^{-1}$$

is also only defined in $[0,1]$. It has

$$P(x) = x^2/\{x^2 + (1-x)^2\}$$

and is not proper, although $P(x) = p$ has a unique root in x for each $0 \leq p \leq 1$ and is single-valued.

As an example of rule with infinite range consider the exponential rule with

$$f(x,1) = e^{-\frac{1}{2}x} \text{ and } f(x,0) = e^{\frac{1}{2}x} .$$

Here $x_F = -\infty$, $x_T = +\infty$ and $P(x) = 1/(1 + e^{-x})$ ranging from 0 to 1. This is improper but nevertheless a possibly useful rule in that it encourages the person to select x corresponding to a probability where $p = 1/(1 + e^{-x})$ and hence $x = \log \{p/(1 - p)\}$. In other words, the values announced are log-odds.

The rules with

$$f(x,1) = 1 - F_1(x) \text{ and } f(x,0) = F_0(x)$$

where $F_i(x)$ are distribution functions on $(-\infty,\infty)$ are interesting because they are bounded both above and below and are defined on $[-\infty,\infty]$. If $f_1(x)$ and $f_0(x)$ are the corresponding densities, $P(x) = f_0(x)/\{f_0(x) + f_1(x)\}$. Often these do not provide acceptable rules since the range of $P(x)$ is not the full unit interval. An extreme case arises with $f_0(x) = f_1(x)$ when $P(x) = \frac{1}{2}$ for all x and only $\pm\infty$ are admissible: see comment (12) below. If $f_1(x)$ corresponds to $N(1,\frac{1}{2})$ and $f_0(x)$ to $N(-1,\frac{1}{2})$, then $P(x) = 1/(1 + e^{-x})$ and we are back to a log-odds rule.

5. The notion of admissibility is essentially that of Pareto optimality. One way of expressing the result of this paper is to say that a person who accepts Pareto optimality and the invariance assumption, and who then, by some unstated process, selects a unique value from the Pareto set, is effectively introducing probabilities and minimizing an expected value. In situations where the single-valued condition does not obtain, many of the values in the Pareto set are ruled out. (Nau's quartic rule illustrates this.)

6. The considerations of this paper have considerable practical import besides the justification of the Bayesian argument.

Consider a geologist who, after a survey, is asked to express his uncertainty about E, the existence of oil at a site, conditional on the result F of the survey. Then he may well see the position in terms of implicit score functions reflecting the dangers of giving a high value, so encouraging drilling, when the area is dry; and the lesser dangers of giving a low value when subsequent drilling reveals oil. It would not be unreasonable to expect that the implicit score function was improper and that he will therefore be motivated to give x rather than his probability p. This suggests that in many cases attention should be paid to the score function so that the stated value may be transformed onto the probability scale. If the geologist provides several assessments then information about the transform, and hence about the scores, can be found from the known probability structure of the transformed value.

It may, of course, happen that the implicit score function just referred to does not obey the regularity conditions. In which case the geologist will be led to make emphatic statements about the existence of oil, as was mentioned in comment 4.

7. We now consider ways of assigning numbers to uncertain events that have been suggested in the literature to see if they lead to admissible values when judged by <u>any</u> scoring rule. For a real parameter θ, the method of (one-sided) confidence intervals enables a number to be attached to the event E, that $\theta < a$, conditional on F, the data: this is the <u>confidence</u> that $\theta < a$ and we write $cf(\theta < a \mid data)$. Suppose

$$cf(\theta < -1 \mid data) = \alpha , \qquad cf(\theta < +1 \mid data) = \beta \tag{3}$$

and

$$cf(\theta < -1 \mid data, \theta < +1) = \gamma . \quad (4)$$

Then, if the confidence method is admissible, we must be able to find $P(\cdot)$ such that $P(\alpha) = P(\beta)P(\gamma)$: this follows from Lemma 3. But since the confidence statement (3) is derived from a probability statement valid for all θ , the restriction to $\theta < +1$ in (4) makes no difference to the validity of the statement and hence $\gamma = \alpha$. Consequently $P(\alpha) = P(\beta)P(\alpha)$ and either $P(\beta) = 1$ or $P(\alpha) = 0$. Hence there is no transform of a confidence statement to a probability statement and the confidence values are inadmissible.

8. Another way of assigning numbers is through significance tests. Let data x have an exponential distribution with density $\theta e^{-\theta x}$, $x \geq 0$, $\theta > 0$. To test the hypothesis that $\theta = \omega$, against the alternative $\theta \neq \omega$, when x is unexpectedly large on hypothesis ω , the "tail" of the null distribution is used:

$$P(X > x \mid \omega) = \int_x^\infty \omega e^{-\omega t} dt = e^{-\omega x}$$

and

$$sg(\omega \mid x) = e^{-\omega x}$$

is the significance attached to the event E that $\theta = \omega$, given F , the data. If x is small, the other tail is used and $sg(\omega \mid x) = 1 - e^{-\omega x}$. Hence for all x

$$sg(\omega \mid x) = \min \{e^{-\omega x} , 1 - e^{-\omega x}\} .$$

For this to correspond to a scoring rule, there must exist a transform $P(\cdot)$ of these values to nonnegative values with the integral over all ω equal to unity: this is the addition rule of probability. But the significance value depends only on ωx, so $\int P(\omega x)d\omega = 1$. Let $\omega x = u$, then $\int P(u)du = x^{-1}$ for all x, which is impossible. Hence significance statements are inadmissible.

9. The discussion in (7) and (8) of confidence and significance statements is based on my personal understanding of these methods. That understanding may be defective because the methods are not unambiguously described. For example, in (7), is the result that led to $\gamma = \alpha$ correct? Is a confidence statement altered if the parametric range is restricted? The discussion of significance levels in (8) is similarly bedevilled by the ambiguity over whether one-or two-sided tests are appropriate: we have used only the one-sided form. It is my conviction that both these methods are inadmissible because they violate the likelihood principle, that easily follows from the probabilistic description of uncertainty.

10. Another way of assigning numbers to uncertain events has been suggested by Zadeh (1979). These numbers are called possibilities. Let all statements be conditional on the same event not described in the notation. Then the possibilities $\Pi(E)$ for events E satisfy the rule of combination

$$\Pi(E \cup F) = \max \{\Pi(E), \Pi(F)\} .$$

This is in conflict with the corresponding probability rule

$$p(E \cup F) = p(E) + p(F) - p(E \cap F)$$

which is a linear operator on the statement for indicator variables

$$1 - (1 - E)(1 - F) = E + F - EF .$$

The possibility relation being nonlinear cannot be transformed and hence possibilities are inadmissible.

11. An extension of the idea of using a single number to describe the uncertainty of E conditional on F is to use two values, x_1 , x_2 . They are sometimes called upper and lower probabilities. To score these, one might use a function $f(x_1,x_2,E)F$. Consider applying the admissibility ideas here. (We omit the details which parallel those given above.)

With (x_1,x_2) stated for E conditional on F , the scores are

$$\left.\begin{array}{ll} EF = 1 & f(x_1,x_2,1) \\ (1 - E)F = 1 & f(x_1,x_2,0) . \end{array}\right\}$$

As before consider small changes δx_1 , δx_2 in the values. Then the score changes will be

$$\left.\begin{array}{l} f_1(x_1,x_2,1)\delta x_1 + f_2(x_1,x_2,1)\delta x_2 \\ \text{and} \quad f_1(x_1,x_2,0)\delta x_1 + f_2(x_1,x_2,0)\delta x_2 \end{array}\right\}$$

where f_i denotes the derivative with respect to the i^{th} argument. For admissibility the determinant must vanish. This determinant is equal to the Jacobian of the transformation from (x_1,x_2) to (f,g) . If it vanishes everywhere, the functions f and g assume constant values on the same curve in the (x_1,x_2)-plane, so that there is no reason to choose between different values on the curve and the subject is effectively only using one number (that describes which curve), rather than two, to measure his uncertainty.

If the Jacobian does not vanish everywhere then the values of (x_1,x_2) are confined to the curve where it does vanish, namely where

$$\frac{f_1(x_1,x_2,1)}{f_2(x_1,x_2,1)} = \frac{f_1(x_1,x_2,0)}{f_2(x_1,x_2,0)} . \tag{5}$$

Call this common value $h(x_1,x_2)$. Then again, in effect, the subject is only providing a single number describing his position on that curve. For example, suppose (x_1,x_2) is given for E, and (y_1,y_2) for $\bar{E}$, both conditional on F. This is the situation comparable to that in Lemma 2 and the total scores are

$$\left.\begin{array}{lll} & EF = 1 & f(x_1,x_2,1) + f(y_1,y_2,0) \\ \text{and} & (1 - E)F = 1 & f(x_1,x_2,0) + f(y_1,y_2,1) . \end{array}\right\}$$

The changes in scores, resulting from changes $(\delta x_1,\delta x_2)$ in (x_1,x_2) and $(\delta y_1,\delta y_2)$ in (y_1,y_2), will be on utilizing (5),

$$f_2(x_1,x_2,1)[h(x_1,x_2)\delta x_1 + \delta x_2] + f_2(y_1,y_2,0)[h(y_1,y_2)\delta y_1 + \delta y_2]$$
$$f_2(x_1,x_2,0)[h(x_1,x_2)\delta x_1 + \delta x_2] + f_2(y_1,y_2,1)[h(y_1,y_2)\delta y_1 + \delta y_2] .$$

The vanishing of the determinant gives

$$f_2(x_1,x_2,1)f_2(y_1,y_2,1) = f_2(x_1,x_2,0)f_2(y_1,y_2,0) ,$$

or if

$$P(x_1,x_2) = \frac{f_2(x_1,x_2,0)}{f_2(x_1,x_2,0) - f_2(x_1,x_2,1)} ,$$

that $P(x_1,x_2) + P(y_1,y_2) = 1$ and we are back to the addition rule for probabilities. The product rule follows similarly.

This does not close the book on the idea of using two or more numbers to describe uncertainty, for it might be reasonable to use two or more score functions, measuring different qualities of the descriptions in the manner of a multiattribute utility function.

12. The argument of Shafer (1976) is affected by the scoring-rule criterion. He suggests, in the situation of Lemma 2, that <u>any</u> values, x for E, y for $\bar{E}$, could be used subject only to the requirements that $x \geq 0$, $y \geq 0$, $x + y \leq 1$. Such numbers are possible values for a belief function. But Lemma 2 shows that $P(x) + P(y) = 1$ and hence the only scoring rule to make all Shafer's values admissible has $P(x) = \frac{1}{2}$, or $f(x,0) + f(x,1) = \text{constant}$. But this contradicts the product rule in Lemma 3. Alternatively $P(x) = \frac{1}{2}$ means that $f'(x,0) = -f'(x,1)$ and hence $\lim_{x \to x_F} f'(x,0)/f'(x,1) = -1$ in contradiction of the regularity condition.

13. Notice that in the score assumption we have supposed that n, the number of events judged, is <u>finite</u>. The infinite case causes difficulties due to the possible divergence of the series describing the total scores. As a result we have only established the addition rule for a finite number of events and the resulting probability is only finitely-additive and not σ-additive. We have been unable to see how, or even if it is possible, to extend the notion of a score to an enumerable infinity of statements.

REFERENCES

DE FINETTI, B. (1974), Theory of Probability, Vol. 1, Wiley, New York.

DEMPSTER, A. P. (1968), "A generalization of Bayesian inference", J. Roy. Statist. Soc. B, 30, 205-247.

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

SHAFER, GLENN (1976), A Mathematical Theory of Evidence, Princeton University Press.

SMITH, CEDRIC A. B. (1961), "Consistency in statistical inference and decision", J. Roy. Statist. Soc. B, 23, 1-37.

ZADEH, L. A. (1979), "Possibility theory and soft data analysis", Memo. UCB/ERL M79/66. University of California, Berkeley.