Mathematical Social Sciences 19 (1990) 159–165
North-Holland

DIAGNOSTIC CONDITIONALIZATION

Carl WAGNER

Department of Mathematics, University of Tennessee, Knoxville, TN 37996-1300, U.S.A.

Bruce TONN

Oak Ridge National Laboratory, Oak Ridge, TN 37831-6207, U.S.A.

Communicated by F.R. McMorris
Received 21 March 1989
Revised 21 July 1989

Dedicated to our friend and colleague Ram Uppuluri

We describe a generalization of ordinary conditionalization with particular relevance to diagnostic problems. Current diagnostic evidence is represented by a type of lower probability, called a belief function. This belief function is then upgraded to a probability measure by means of an empirical probability recording past relative frequencies of the relevant diagnostic categories.

Key words: Conditionalization; diagnostic; belief function.

1. Updating versus upgrading

Conditionalization is commonly thought of as the updating of a prior probability in the light of new evidence. In the simplest example, the prior p on X is (given evidence which renders certain some $E \subseteq X$) updated to $q(\cdot) = p(\cdot \mid E)$. Generalizations of this technique have been developed for use with less decisive sorts of evidence (see Jeffrey, 1965). Whatever their degree of generality, however, underlying all these procedures is the view that one discards the prior p in favor of a new and improved q. While this view comports well with the idea of the truth-seeker-as-juror (forming initial impressions, refining them in the face of new evidence), it is deficient as a model of the truth-seeker-as-diagnostician. In what follows, we shall argue that the latter sort of truth-seeker is best understood as combining diagnostic evidence to construct a type of lower probability known as a belief function (see Shafer, 1976, and Section 2 below) and then, if necessary, upgrading that belief function to a probability measure by drawing on an empirically based probability p that records past relative frequencies of the relevant diagnostic categories. Far from being discarded after this procedure, p remains a valuable part of the knowledge base, to be drawn on as required in future diagnostic exercises.

2. Belief functions

A *belief function* (see Shafer, 1976) on the finite set X is a mapping $b: 2^X \to [0,1]$ such that $b(\varnothing)=0$, $b(X)=1$, and for all positive integers r and every sequence $A_1, \ldots, A_r$ of subsets of X:

$$b(A_1 \cup \cdots \cup A_r) \geq \sum_{\substack{I \subseteq \{1,\ldots,r\} \\ I \neq \varnothing}} (-1)^{|I|-1} b\Big(\bigcap_{i \in I} A_i\Big). \tag{2.1}$$

In particular, $b(A_1)+b(A_2) \leq b(A_1 \cup A_2)$ if $A_1 \cap A_2 = \varnothing$, and so $b(A)+b(\bar{A}) \leq b(X)=1$. That belief functions need only be superadditive makes them attractive measures of uncertainty, for one can honestly represent the case where one has little evidence for or against A by assigning both A and $\bar{A}$ small (even zero) b-measures. Every probability measure is, of course, a belief function, since (2.1) is implied by the principle of inclusion and exclusion for probabilities. Indeed, a belief function b is a probability measure if and only if $b(A)+b(\bar{A})=1$ for all $A \subseteq X$ (see Shafer, 1976, Theorem 2.8).

Among belief functions that are not probability measures the simplest is the *trivial belief function* b_0, which assigns every proper subset of X belief zero (representing the case where all one knows is that the truth lies somewhere in X). Only slightly more complex are the *simple support functions* $b_{E;s}$, defined for each $E \subseteq X$ and $s \in [0,1]$ by $b_{E;s}(A)=s$ if $E \subseteq A \neq X$, $b_{E;s}(X)=1$, and $b_{E;s}(A)=0$ otherwise. Evidence which renders E certain (the context of classical conditionalization) is represented by the simple support function $b_{E;1}$.

It is not entirely trivial to verify that b_0 and $b_{E;s}$ satisfy (2.1). Indeed, the direct construction of a mapping $b: 2^X \to [0,1]$, and the verification that it satisfies (2.1) is in general a difficult task. This task is facilitated by the existence of a class of auxiliary mappings $m: 2^X \to [0,1]$, called *basic probability assignments* (BPAs), defined by the properties $m(\varnothing)=0$ and

$$\sum_{A \subseteq X} m(A) = 1. \tag{2.2}$$

Every BPA m on X induces a belief function $b^{(m)}$ by the formula:

$$b^{(m)}(A) = \sum_{E \subseteq A} m(E), \tag{2.3}$$

and every belief function b on X induces a BPA $m^{(b)}$ by the formula:

$$m^{(b)}(A) = \sum_{E \subseteq A} (-1)^{|A-E|} b(E), \tag{2.4}$$

with $m^{(b^{(m)})}=m$ and $b^{(m^{(b)})}=b$ (see Shafer, 1976, pp. 38–40). One can thus use BPAs either to construct[1] belief functions [by (2.3)] or to verify that an uncertainty measure $b: 2^X \to [0,1]$ is a belief function [by checking that the quantities $m^{(b)}(A)$

[1] See Shafer (1976, 1981) for a description of the sorts of introspection involved in the assessment of BPAs and belief functions.

defined by (2.4) are non-negative].[2] In addition, BPAs provide a criterion for b to be a probability measure, for given a belief function b, and defining the *focal elements* of b to be those $E \subseteq X$ such that $m^{(b)}(E) > 0$, it follows that b is a probability measure if and only if each of its focal elements is a singleton (see Shafer, 1976, Theorem 2.8).

In addition to arising by direct assessment (mediated or unmediated by some BPA) belief functions also arise indirectly in the form of certain lower probabilities, first studied by Dempster (1967). Specifically, suppose that evidence enables us to assess on a related possibility set Y a probability mass function u, and that we understand the relation between outcomes in Y and those in X sufficiently to define a mapping $C: Y \to 2^X - \{\varnothing\}$, with $C(y)$ construed as the set of outcomes in X not precluded by the outcome $y \in Y$.[3] It follows that the mapping $b_{u;C}: 2^X \to [0, 1]$, defined by

$$b_{u;C}(E) = \sum_{\substack{y \in Y: \\ C(y) \subseteq E}} u(y), \tag{2.5}$$

is a belief function on X, since the mapping $m_{u;C}: 2^X \to [0, 1]$, defined by

$$m_{u;C}(E) = \sum_{\substack{y \in Y: \\ C(y) = E}} u(y), \tag{2.6}$$

is clearly a BPA, and $b_{u;C} = b^{(m_{u;C})}$, as defined by (2.3). As one would expect, $b_{u;C}$ is a probability measure precisely when $C(y)$ is a singleton for every $y \in Y$ (so that, in essence, C is an X-valued random variable). Since any way of assigning a probability measure to a set $E \subseteq X$ should render E at least as probable as the set of outcomes in Y that preclude all outcomes outside E, it follows from (2.5) that any probability measure q subsequently materializing on X ought to be bounded below by $b_{u;C}$. For this reason, the mapping $b_{u;C}$ may be thought of as a type of lower probability.

3. Upgrading belief functions to probability measures

The lower probabilities furnished by the values of a belief function b may well

[2] One need not check that these quantities sum to one. They always do, since

$$\begin{aligned}\sum_{A \subseteq X} m^{(b)}(A) &= \sum_{A \subseteq X} \sum_{E \subseteq A} (-1)^{|A-E|} b(E) \\ &= \sum_{E \subseteq X} b(E) \sum_{X \supseteq A \supseteq E} (-1)^{|A-E|} \\ &= \sum_{E \subseteq X} b(E) \sum_{k=0}^{|X-E|} \binom{|X-E|}{k} (-1)^k = \sum_{E \subset X} b(E) \cdot 0 + b(X) = 1.\end{aligned}$$

[3] We assume that no $y \in Y$ precludes every $x \in X$, in the interests of simplicity. Dempster (1967) actually deals with the more general case of $C: Y \to 2^X$. As anyone who has been told by a physician, 'I can tell you what it isn't', knows, the specification of C functions is a common diagnostic exercise.

suffice for decision-making purposes. Suppose, however, that it is desirable to attain the more precise assessments of uncertainty furnished by a probability measure. In general, there will be an infinite number of probability measures q on X compatible with b in the sense that $q(E) \geq b(E)$ for all $E \subseteq X$. Let us denote by $P(X; b)$ the set of all such compatible probability measures. The set $P(X; b)$ is convex (i.e. closed under weighted arithmetic averaging of any finite set of its members) and, unless b is itself a probability measure [in which case $P(X; b) = \{b\}$], infinite.[4] The members of $P(X; b)$ may be characterized quite concretely: they are simply all of the probability measures which arise by allocating, in all possible ways, the associated BPA values $m^{(b)}(E)$ among the individual elements of E, for each $E \subseteq X$.

Theorem. *Let b be a belief function on X, with associated BPA $m^{(b)}$ defined by (2.4). A probability measure q on X is compatible with b if and only if, for every $E \subseteq X$, there exists a function $s_E : X \to [0, 1]$, with $s_E(x) = 0$ if $x \notin E$ and $\sum_x s_E(x) = m^{(b)}(E)$, such that for all $x \in X$:*

$$q(x) = \sum_{E \subseteq X} s_E(x). \tag{3.1}$$

Proof. Dempster (1967) has established this result for belief functions of the types defined by formula (2.5). But every belief function b on X arises, abstractly, in such a fashion. For given a belief function b on X, let $Y = \{E \subseteq X: m^{(b)}(E) > 0\}$, let $u(E) = m^{(b)}(E)$ for all $E \in Y$, and let C: $Y \to 2^X - \{\varnothing\}$ by $C(E) = E$. Clearly, $m^{(b)} = m_{u;\,C}$, as defined by (2.6), and so $b = b_{u;\,C}$, as defined by (2.5). Hence the theorem holds for all belief functions.

There is an instructive equivalent formulation of (3.1), based on the observation that one can ignore those E for which $m^{(b)}(E) = 0$, and on the additivity of q, namely that for all $A \subseteq X$:

$$q(A) = \sum_{E \in \mathscr{E}} \sum_{x \in A} s_E(x), \tag{3.2}$$

where

$$\mathscr{E} = \{E \subseteq X: m^{(b)}(E) > 0\}, \tag{3.3}$$

the family of focal elements of b.

Choosing from $P(X; b)$ a particular probability measure q to which to upgrade b amounts, as we have seen, to choosing a family $\{s_E: E \in \mathscr{E}\}$ of allocation functions indexed on the focal elements of b. The variety of upgrading problems and the differing amounts of supplementary information that might be available make it doubtful that any universal strategy for choosing such a family can be defended.

[4] That $P(X; b)$ is infinite when b is not a probability measure follows from the aforementioned fact that in such a case there exists at least one nonsingleton E such that $m^{(b)}(E) > 0$, and from the theorem below.

We wish to mention, however, a strategy with what we believe to be fairly broad applicability. This strategy depends on the availability of an empirically based probability p on X (not an unrealistic assumption in many diagnostic contexts) and defines the relevant allocation functions s_E by

$$s_E(x) = p(\{x\} \mid E)m^{(b)}(E),^5 \tag{3.4}$$

for each $E \in \mathscr{E}$. With this proportional-to-p allocation, formula (3.2) takes the elegant form:

$$q(A) = \sum_{E \in \mathscr{E}} m^{(b)}(E)p(A \mid E). \tag{3.5}$$

To adopt (3.4), and hence (3.5), is to judge that our diagnostic evidence is exhausted in the assessment of b (equivalently, $m^{(b)}$) and that, otherwise, the uncertainty assessments based on relative frequencies of past diagnoses are reasonably assumed to remain operative. [See Wagner, 1989, for further discussion of this issue, including a specification of conditions entailing (3.4) when b is of the form (2.5).]

We remark that when the members of the set of focal elements $\mathscr{E}$ are pairwise disjoint, then $q(E) = m^{(b)}(E) = b(E)$ for all $E \in \mathscr{E}$ and so (3.5) yields:

$$q(A) = \sum_{E \in \mathscr{E}} q(E)p(A \mid E), \tag{3.6}$$

a well known conditionalization rule first explored by Jeffrey (1965).

Example. The microwave communication system. In a certain microwave communication system,[6] alarm evidence of uncertain quality is to be used to assess the mutually exclusive, exhaustive possibilities of a station problem (s), a transmitter problem (t), a receiver problem (r), or a false alarm (f). Of relevance to the set $X = \{s, t, r, f\}$ of possible diagnoses is the set $Y = \{(a_1, a_2, a_3, a_4)\}: a_i = 0$ or 1, and $a_2 a_3 = 0\}$, with a given quadruple in Y indicating the presence ($a_i = 1$) or absence ($a_i = 0$) of each of four alarm-based conditions $i = 1$, 2, 3, and 4, conditions 2 and 3 being, as a matter of definition, incompatible. The mapping $C: Y \to 2^X - \{\varnothing\}$, given by Table 1, associates to each quadruple in Y the set of diagnoses not precluded by the pattern of alarm-based conditions defined by that quadruple.

Completely trustworthy alarm evidence would establish with certainty a single quadruple y_i from the list in Table 1, which would in turn induce on X the simple support function $b_{C(y_i);1}$ described in Section 2. In practice, however, alarm evidence is unlikely to be so reliable. In particular, the trustworthiness of an alarm may decrease as the time elapsed since the initial receipt of the alarm increases. Let us suppose that such considerations have been taken into account in the form of a

[5] It is implicit in (3.4) that p is positive on every focal element.

[6] We are grateful to Tom Wiggen for furnishing us with this example. See also Goeltz, MacGregor, Purucker, Tonn and Wiggen (1989).

Table 1

$y_i = (a_1, a_2, a_3, a_4)$	$C(y_i)$
$y_1 = (1, 1, 0, 1)$	X
$y_2 = (1, 1, 0, 0)$	$\{s, t, f\}$
$y_3 = (1, 0, 1, 1)$	X
$y_4 = (1, 0, 1, 0)$	X
$y_5 = (1, 0, 0, 1)$	X
$y_6 = (1, 0, 0, 0)$	$\{s, f\}$
$y_7 = (0, 1, 0, 1)$	$\{t, r, f\}$
$y_8 = (0, 1, 0, 0)$	$\{t, f\}$
$y_9 = (0, 0, 1, 1)$	$\{t, r, f\}$
$y_{10} = (0, 0, 1, 0)$	$\{t, r, f\}$
$y_{11} = (0, 0, 0, 1)$	$\{t, r, f\}$
$y_{12} = (0, 0, 0, 0)$	$\{f\}$

probability mass function u on Y defined by $u(y_i)=0$ for $1 \le i \le 6$, $u(y_7)=0.010$, $u(y_8)=0.002$, $u(y_9)=0.325$, $u(y_{10})=0.080$, $u(y_{11})=0.468$, and $u(y_{12})=0.115$.

From (2.6), with u and C as specified above, it follows that the nonzero values of the BPA $m=m_{u;C}$ are given by $m(E_1)=0.002$, $m(E_2)=0.883$, and $m(E_3)=0.115$, where $E_1=\{t,f\}$, $E_2=\{t,r,f\}$ and $E_3=\{f\}$. Corresponding values of the belief function $b=b_{u;C}$ are easily calculated.

To upgrade b to a compatible probability q on X requires a choice of allocation functions s_{Ei}, $i=1, 2, 3$, to be incorporated in formula (3.2). Let us consider an example of such a choice, the case in which the past relative frequency p of diagnoses in X is known, say $p(s)=0.20$, $p(t)=0.30$, $p(r)=0.35$, and $p(f)=0.15$, and the proportional-to-p allocation represented by formula (3.5) is adopted. Then, for all $A \subseteq X$, $q(A)=(0.002)p(A \mid E_1)+(0.883)p(A \mid E_2)+(0.115)p(A \mid E_3)$. In particular, $q(s)=0$, $q(t)=0.332$, $q(r)=0.387$, and $q(f)=0.281$.

4. Other approaches to upgrading

We have proposed that the proper way to upgrade a belief function b to a probability measure is to make a considered choice of a family of allocation functions $\{s_E\}$ and to upgrade b to the probability measure q defined, equivalently, by (3.1) or (3.2). We also described one possible strategy for choosing the functions s_E, given the availability of a relevant empirically based probability p on the possibility set X.

There is no disputing that the choice of a family of allocation functions may be a demanding intellectual exercise. Consequently, it may be tempting to opt for a 'mechanical'[7] approach to upgrading b. We have in mind here such strategies as upgrading b to a probability measure $q \in P(X; b)$ with maximum entropy $H(q) = -\sum q(x)\log(q(x))$. A variant on this strategy, given a probability p on X, as describ-

[7] We have adopted this term from Diaconis and Zabell (1982), who use it in a narrower context.

ed in the previous paragraph, would be to upgrade b to a probability measure $q \in P(X; b)$ which minimizes the 'distance' from p as measured by the Kullback–Leibler number $I(q, p) = \sum q(x)\log(q(x)/p(x))$. Apart from the difficulty of rationalizing the choice of precisely this measure of the distance from q to p (there are many other possible measures, several of which are described in Diaconis and Zabell, 1982) there are major objections to so-called MAXENT strategies, which have been clearly and (to us) convincingly articulated by Skyrms (1987). Among these objections is the fact that when the aforementioned probability p happens itself to be compatible with b, a MAXENT strategy will upgrade to p, in effect ignoring current diagnostic evidence as incorporated in b, and opting to rely on background statistics.

An entirely different (but no less deficient) strategy for upgrading b in the presense of p would be to upgrade b to $q = p * b$, the result of combining p and b by Dempster's rule of combination (see Shafer, 1976, Chapter 3). While q, so defined, is always a probability measure, it may well fail to meet the basic requirement that it be compatible with (that is, bounded below by) b.[8] Dempster's rule may, of course, play an important role in the construction of b itself, by building up b from simpler belief functions (like the simple support functions mentioned in Section 2) based on distinct bodies of diagnostic evidence. But its attractiveness as a rule of combination should not mislead one into thinking that it can function generally as a rule of conditionalization.

[8] To take a very simple example, suppose that $X = \{x_1 x_2\}$, $p(x_1) = 0.4$, $p(x_2) = 0.6$ and b is actually a probability measure on X, say $b(x_1) = 0.2$ and $b(x_2) = 0.8$. By Dempster's rule (see Shafer, 1976, Chapter 3):

$$p * b(x_1) = (0.2)(0.4)/1 - (0.2)(0.4) - (0.8)(0.6) \approx 0.18.$$

On the other hand, any upgrading formula of type (3.2), including (3.5), will yield $q = b$. This is eminently reasonable. If a patient's symptoms alone enable a physician to assess a probability measure over the possible ailments causing these symptoms, the incidence of those ailments in the general population is no longer relevant.

References

A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.

P. Diaconis and S. Zabell, Updating subjective probability, J. Amer. Stat. Assoc. 77 (1982) 822–830.

R. Goeltz, D. MacGregor, S. Purucker, B. Tonn and T. Wiggen, Bonneville Power Administration communication alarm processor expert system: phase I description, Oak Ridge National Laboratory Report, 1989.

R. Jeffrey, The Logic of Decision (McGraw-Hill, New York, 1965).

G. Shafer, A Mathematical Theory of Evidence (Princeton University Press, Princeton, 1976).

G. Shafer, Constructive probability, Synthese 48 (1981) 1–60.

B. Skyrms, Updating, supposing, and maxent, Theory and Decision 22 (1987) 225–246.

C. Wagner, A natural generalization of Jeffrey conditionalization, Technical Report, University of Tennessee, Knoxville, TN, 1989.