DEMPSTER'S RULE OF COMBINATION

by

Glenn Shafer

University of Kansas

Summary. Dempster's Rule of Combination.

The theory of belief functions is a generalization of probability theory; a belief function is a set function more general than a probability measure but whose values can still be interpreted as degrees of belief. Dempster's rule of combination is a rule for combining two or more belief functions; when the belief functions combined are based on distinct or "independent" sources of evidence, the rule corresponds intuitively to the pooling of evidence. As a special case, the rule yields a rule of conditioning which generalizes the usual rule for conditioning probability measures. The rule of combination was studied extensively, but only in the case of finite sets of possibilities, in the author's monograph A Mathematical Theory of Evidence. The present paper describes the rule for general, possibly infinite, sets of possibilities. We show that the rule preserves the regularity conditions of continuity and condensability, and we investigate the two distinct generalizations of probabilistic independence which the rule suggests.

AMS 1970 subject Classifications. Primary 60A05; Secondary 62A99.

Key words and phrases. Belief function, upper probabilities, allocation of probability, conditioning, Dempster's rule, product measure, cognitive independence, evidential independence, continuity, condensability, alternating set function, monotone set function.

1 Some of the work in this paper was done while the author was supported by a National Science Foundation Graduate Fellowship. The work was completed with support from the General Research Fund of the University of Kansas, allocations #3991-x038 and #3315-x038.
§1. Introduction.

A function $f$ defined on a power set $\mathcal{P}(\Omega)$ is called a belief function if $f(\emptyset) = 0$, $f(\Omega) = 1$, and $f$ is a monotone of order $\infty$. The author's monograph *A Mathematical Theory of Evidence* (1976) studies belief functions in detail under the assumption that $\Omega$ is finite. The resulting theory is centered on Dempster's rule of combination, a rule for combining two belief functions $f_1$ and $f_2$, both defined on the same power set $\mathcal{P}(\Omega)$, to obtain their orthogonal sum $f_1 \oplus f_2$, which is also a belief function on $\mathcal{P}(\Omega)$. This rule is central because of its intuitive interpretation: if $f_1$ and $f_2$ express degrees of belief based on entirely distinct bodies of evidence, then the operation of forming $f_1 \oplus f_2$ is interpreted as pooling the two bodies of evidence.

The present paper, which studies Dempster's rule for an arbitrary, possibly infinite set $\Omega$, is a sequel to Shafer (1979), which studies the representation and extension of belief functions on infinite sets. We freely use the vocabulary, notation and results of that paper.

In preparation for describing Dempster's rule (§6 below), we adduce two special cases: the rule for forming product belief functions (§3) and the rule of conditioning (§5). Both these rules generalize the corresponding rules for probability measures, and both preserve the regularity conditions of continuity and condensability. Like the proof that the product of two countably additive measures is countably additive, the proof that the product of two continuous belief functions is continuous requires a notion of integration—in this case a notion of upper integration, which is explained in §2.
In §§4 and 7 we generalize to the case of infinite sets of possibilities the notions of evidential and cognitive independence of subalgebras with respect to a belief function. Here, as in the finite case (see Chapter 7 of Shafer (1976a)), evidential independence requires that both the belief function and its upper probability function obey the usual rule
\[ P(A \cap B) = P(A)P(B), \]
while cognitive independence requires only that the upper probability function obey this rule. The notion of evidential independence is intuitively retrospective: two subalgebras are evidentially independent with respect to a belief function if that belief function can be obtained by combining a belief function that bears on only one of them with a belief function that bears on only the other. The weaker notion of cognitive independence, on the other hand, is intuitively prospective: two subalgebras are cognitively independent with respect to a belief function if combination with a new belief function that bears on only one of them does not change the degrees of belief for elements of the other.

It should be noted that the study of monotone and alternating set functions was initiated by Gustave Choquet (1953); and that the theory of belief functions is closely related to Choquet's work; see §2 of Shafer (1979). The theory of belief functions derives, however, from work by A. P. Dempster (1967, 1968), work which was independent of Choquet's work and was directed towards problems in statistical inference. Dempster used the name "lower probabilities" for the quantities here called "degrees of belief," and he introduced his lower and upper probabilities by means of a multivalued mapping from a measure space. He also described his rule of combination in terms of such mappings: see Dempster (1967). The axiomatic approach taken here and in Shafer
(1976a, 1979 ) permits a rigorous study of the rule in connection with
the notions of continuity and condensability.

Some of the results in this paper were originally obtained in
the author's doctoral dissertation (1973); some were announced in
Shafer (1976b). See also Nguyen (1978).

§2. Lower and Upper Integration.

Suppose $g$ is a bounded function on $P(\Omega)$ such that $g(\emptyset) = 0$
and $g(A) \leq g(B)$ whenever $A \subset B \subset \Omega$. Given an extended real-valued
function $\varphi$ on $\Omega$, we set

$$G(\varphi) = \int_0^\infty g(\varphi^{-1}(x, \infty]) dx - \int_{-\infty}^0 (g(\Omega) - g(\varphi^{-1}(x, \infty])) dx$$

(2.1)

whenever at least one of the integrals on the right-hand side of this
equation is finite. When $\varphi$ is non-negative, this reduces, of course,
to

$$G(\varphi) = \int_0^\infty g(\varphi^{-1}(x, \infty]) dx .$$

(2.2)

This functional has been studied extensively, though usually with an
emphasis on some topology for $\Omega$; see, for example, Choquet (1953,
p. 265) and Huber and Strassen (1973).

Notice that the set $(x, \infty]$ could be replaced, in (2.1) or (2.2),
by the set $[x, \infty]$. For this can change the integrands only at their
points of discontinuity, and since they are monotonic, there are only
a countable number of these. Notice also that if $A \subset \Omega$ and $\chi_A$
is its characteristic function, then $G(\chi_A) = g(A)$; hence the functional
$G$ may be thought of as an extension of the set function $g$. Moreover,
if $g$ agrees with a measure $\mu$ on some $\sigma$-algebra $\mathcal{G}$ of subsets of $\Omega$
and \( \varphi \) is measurable with respect to \( G \), then \( G(\varphi) \) is the integral of \( \varphi \) with respect to \( \mu \).

We shall be mainly concerned with the case where \( \varphi \) is non-negative. Accordingly, we denote by \( \Phi^+ \) the set of all non-negative extended real-valued functions of \( \Omega \) and note the following facts:

1. If \( \varphi \in \Phi^+ \) and \( a > 0 \) then \( G(a\varphi) = aG(\varphi) \).

2. If \( \varphi_1, \varphi_2 \in \Phi^+ \) and \( \varphi_1 \leq \varphi_2 \), then \( G(\varphi_1) \leq G(\varphi_2) \).

3. If \( g(\bigcap A_i) = \lim_{i \to \infty} g(A_i) \) for every decreasing sequence \( A_1 \supset A_2 \supset \ldots \) in \( \mathcal{P}(\Omega) \), then \( G(\lim_{i \to \infty} \varphi_i) = \lim_{i \to \infty} G(\varphi_i) \) for every decreasing sequence \( \{\varphi_i\} \) in \( \Phi^+ \).

4. If \( g(\bigcup A_i) = \lim_{i \to \infty} g(A_i) \) for every increasing sequence \( A_1 \subset A_2 \subset \ldots \) in \( \mathcal{P}(\Omega) \), then \( G(\lim_{i \to \infty} \varphi_i) = \lim_{i \to \infty} G(\varphi_i) \) for every increasing sequence \( \{\varphi_i\} \) in \( \Phi^+ \).

5. If \( g \) is monotone of order 2, then

\[
G(\varphi_1) + G(\varphi_2) \leq G(\varphi_1 + \varphi_2)
\]

for all \( \varphi_1, \varphi_2 \in \Phi^+ \).

6. If \( g \) is alternating of order 2, then

\[
G(\varphi_1) + G(\varphi_2) \geq G(\varphi_1 + \varphi_2)
\]

for all \( \varphi_1, \varphi_2 \in \Phi^+ \).

The first four of these facts are easy to verify; (5) and (6) were proven by Choquet (1953, pp. 287–288).

A function \( \varphi \in \Phi^+ \) is called a simple function if it can be written in the form

\[
\varphi = \sum_{i=1}^{n} a_i \chi_{A_i}
\]  
(2.3)
where \( n \geq 1 \), each \( A_i \) is a subset of \( \Omega \), \( \chi_{A_i} \) is the characteristic function for \( A_i \), and each \( a_i \) is a non-negative real number. Every simple function \( \varphi \) in \( \Phi^+ \) can be expressed in the form (2.3) with the \( A_i \) nested.

**Theorem 2.1.** Suppose \( \varphi \in \Phi^+ \) is given by (2.3).

(i) If the \( A_i \) are nested, then

\[
G(\varphi) = \sum_{i=1}^{n} a_i g(A_i).
\]

(ii) In general,

\[
G(\varphi) = \sum_{I \subseteq \{1, \ldots, n\}, \text{ } I \neq \emptyset} (\max_{i \in I} a_i) \sum (-1)^{|J|+1} g\left( \bigcup_{i \in J}^{-1} A_i \right); \quad (2.4)
\]

here \( \overline{I} \) denotes the complement of \( I \) with respect to \( \{1, \ldots, n\} \), and we use the convention that \( \bigcup_{i \in \emptyset} A_i = \emptyset \).

**Proof:** (i) We may assume without loss of generality that \( A_1 \supset A_2 \supset \ldots \supset A_n \). And in this case the result follows from the fact that

\[
\varphi^{-1}(x, \infty] = \begin{cases} 
A_1, & 0 \leq x < a_1 \\
A_k, & \sum_{i=1}^{k-1} a_i \leq x < \sum_{i=1}^{k} a_i \\
\emptyset, & \sum_{i=1}^{n} a_i \leq x < \infty.
\end{cases}
\]

(ii) We may assume without loss of generality that the \( a_i \) are ordered: \( 0 \leq a_1 \leq a_2 \leq \ldots \leq a_n \). Set \( b_1 = a_1 \) and \( b_i = a_i - a_{i-1} \) for \( i = 2, \ldots, n \). And set \( B_i = \bigcup_{j=i}^{n} A_j, \quad i = 1, \ldots, n \). Then
\[ \varphi = \sum_{i=1}^{n} a_i \chi_{A_i} = \sum_{i=1}^{n} b_i \chi_{B_i} \]. And since the \( B_i \) are nested,

\[ G(\varphi) = \sum_{i=1}^{n} b_i \ gamma(B_i) = \sum_{i=1}^{n} a_i (\gamma(A_i \cup A_{i+1} \cup \ldots \cup A_n) - \gamma(A_{i+1} \cup \ldots \cup A_n)). \quad (2.5) \]

(Since \( \cup A_i = \emptyset \), the coefficient of \( a_n \) is this last sum is simply \( \gamma(A_n) - \gamma(\emptyset) = \gamma(A_n) \).) But since the \( a_i \) are ordered,

\[ \frac{1}{\sum_{I \subseteq \{1, \ldots, n\}} \sum_{i \in I} (-1)^{|I|} \gamma(\cup \{ A_i \mid i \in J \cup \emptyset \})} \]

\[ = \sum_{i=1}^{n} \sum_{K \subseteq \{1, \ldots, n\}} \gamma(\cup A_j) \sum_{J \subseteq K} (-1)^{|J|} (\gamma(\cup \{ i \} \cap K) + 1) \quad \text{if } K \subseteq \{1, \ldots, i-1\} \cup \{i\} \subseteq K. \quad (2.6) \]

The sum

\[ \sum_{i=1}^{n} \sum_{K \subseteq \{1, \ldots, n\}} \gamma(\cup A_j) \sum_{J \subseteq K} (-1)^{|J|} (\gamma(\cup \{ i \} \setminus K) + 1) \quad \text{if } K \subseteq \{1, \ldots, i-1\} \cup \{i\} \subseteq K \]

is empty unless \( K \supseteq \{i+1, \ldots, n\} \), in which case it becomes

\[ \sum_{i=1}^{n} \sum_{K \subseteq \{1, \ldots, n\}} \gamma(\cup A_j) \sum_{J \subseteq \{1, \ldots, i-1\} \cap K} (-1)^{|J|} (\gamma(\cup \{ i \} \setminus K) + 1) \quad \text{if } K = \{i, i+1, \ldots, n\} \]

\[ = \begin{cases} 
0 & \text{if } \{1, \ldots, i-1\} \cap K \neq \emptyset \\
1 & \text{if } K = \{i, i+1, \ldots, n\} \\
-1 & \text{if } K = \{i+1, \ldots, n\}. 
\end{cases} \]

(2.6) reduces, therefore to (2.5).
Suppose $f$ is a belief function on $P(\Omega)$ and $f^*$ is its upper probability function. Then we can extend both $f$ and $f^*$ to functionals by (2.1); that is to say, we can set

$$f(\varphi) = \int_0^\infty f(\varphi^{-1}(x, \infty])dx - \int_{-\infty}^0 (1-f(\varphi^{-1}(x, \infty]))dx \quad (2.7)$$

and

$$f^*(\varphi) = \int_0^\infty f^*(\varphi^{-1}(x, \infty])dx - \int_{-\infty}^0 (1-f^*(\varphi^{-1}(x, \infty]))dx. \quad (2.8)$$

Notice that $f^*(-\varphi) = -f(-\varphi)$. We call (2.7) and (2.8) the lower and upper integrals, respectively, of $\varphi$ with respect to $f$. We see from (5) and (6) above that these "integrals" are subadditive and superadditive, respectively; if $f$ is continuous, then they are countably so.

Notice that when $f^*$ is an upper probability function with

a lowment $\xi : P(\Omega) \rightarrow M$, formula (2.4) becomes transparent; it says simply that

$$f^*(\sum_{i=1}^n a_i X_i) = \sum_{I \subseteq \{1, \ldots, n\}} \mu(\bigwedge_{i \in I} \xi(A_i) - \bigvee_{i \in I} \xi(A_i)) \quad (2.9)$$

(Here our convention is that $\bigvee_{i \in \emptyset} M_i = \bot$.)

§3. Product Belief Functions.

If given evidence supports a given proposition $A$ to degree $s_1$, and other evidence supports an unrelated proposition $B$ to degree $s_2$, then to what extent does the pooled evidence support the conjunction $A \cap B$? A common-sense answer is $s_1 s_2$. (See §4.4 of Shafer (1976a).) And this simple idea can be applied to whole belief functions.
Suppose, indeed, that \( f_1 \) is a belief function on \( \mathcal{P}(\Omega_1) \) and \( f_2 \) is a belief function on \( \mathcal{P}(\Omega_2) \). Define \( f \) on

\[
\mathcal{E} = \{A \times B \mid A \subseteq \Omega_1; B \subseteq \Omega_2\} \subseteq \mathcal{P}(\Omega_1 \times \Omega_2)
\]

by

\[
f_0(A \times B) = f_1(A)f_2(B).
\]  

(3.1)

**Theorem 3.1.** \( f_0 \) is a belief function.

**Proof:** For \( i = 1, 2 \), let \((M_i, \mu_i)\) be a probability algebra and let

\( r_i : \mathcal{P}(\Omega_i) \rightarrow M_i \)

be an allocation of probability for \( f_i \). Let \( M = M_1 \times M_2 \)
be a product algebra, with product measure \( \mu = \mu_1 \times \mu_2 \). And set

\( \rho_i = h_i \circ r_i \), where \( h_i \) is the canonical homomorphism of \( M_i \) into \( M \). (See Kappos (1969).) Notice that \( \rho_i \) maps \( \mathcal{P}(\Omega_i) \) into \( M_i \), and

\( f_i = \mu^\circ \rho_i \). And

\[
\mu(M_1 \land M_2) = \mu(M_1)\mu(M_2)
\]

(3.2)

whenever \( M_i \) is in the subalgebra of \( M \) generated by \( \rho_i(\mathcal{P}(\Omega_i)) \), \( i = 1, 2 \).

(The relation (3.2) will be used repeatedly in this paper. Notice that
it implies in particular that the two subalgebras are independent: if
\( M_1 \land M_2 = \Lambda \), then either \( M_1 = \Lambda \) or \( M_2 = \Lambda \).

Define \( \rho_o : \mathcal{E} \rightarrow M \) by \( \rho_o(A \times B) = \rho_1(A)\rho_2(B) \). Then \( \rho_o \) is
an allocation of probability, and since \( f_o = \mu^\circ \rho_o \), it follows that \( f_o \)
is a belief function.

\[\square\]

Let us denote by \( f \) the canonical extension of \( f_o \) to the algebra \( \mathcal{G} \)
generated by the multiplicative subclass \( \mathcal{E} \). Intuitively, the partial
beliefs expressed by $f$ are appropriate when the evidence concerning
$\Omega_1 \times \Omega_2$ can be divided into two distinct and totally independent bodies
of evidence, one of which is relevant only to $\Omega_1$ and corresponds to
$f_1$, and the other of which is relevant only to $\Omega_2$ and corresponds
to $f_2$.

Since the domain of $f$ is an algebra, the domain of $f^*$ is this
same algebra. And, quite remarkably, $f^*$ obeys the same multiplicative
rule as $f$.

**Theorem 3.2.** If $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$, then

$$f(A \times B) = f_1(A)f_2(B)$$

and

$$f^*(A \times B) = f_1(A)f_2(B)$$

**Proof:** $f$ is overtly defined to make (3.3) true. To prove (3.4),
consider the canonical extension of $\rho_o$ to $G$, which we denote by
$\rho$. We have

$$\rho(C) = \vee \{ \rho_1(A_1) \wedge \rho_2(A_2) \mid A_1 \subseteq \Omega_1; A_2 \subseteq \Omega_2; A_1 \times A_2 \subseteq C \}.$$  \hspace{1cm} (3.5)

for all $C \in G$. If we fix $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$, the another pair of sets
$A_1 \subseteq \Omega_1$ and $A_2 \subseteq \Omega_2$ will satisfy $A_1 \times A_2 \subseteq \overline{A} \times \overline{B}$ if and only if
$A_1 \subseteq \overline{A}$ or $A_2 \subseteq \overline{B}$. Hence (3.5) yields

$$\rho(\overline{A} \times \overline{B}) = \rho_1(\overline{A}) \vee \rho_2(\overline{B}).$$

This is equivalent to

$$\zeta(A \times B) = \zeta_1(A) \wedge \zeta_2(B),$$

(3.6)
where $\zeta$, $\zeta_1$ and $\zeta_2$ denote the allowances for $\rho, \rho_1$ and $\rho_2$, respectively; (3.4) follows by (3.2).

By evaluating the measure of the right-hand side of (3.5), one can obtain an explicit though unwieldy formula for $f$ in terms of $f_1$ and $f_2$. One can also easily derive from (3.5) the corresponding formula for $\zeta$:

$$\zeta(A) = \wedge\{\zeta_1(A_1) \vee \zeta_2(A_2) \mid A_1 \subset \Omega_1; A_2 \subset \Omega_2; A = (A_1 \times \Omega_2) \cup (\Omega_1 \times A_2)\}. \tag{3.7}$$

Notice also that these formulae remain valid for the extensions of $\rho_o$ and $f_o$ to all of $\mathcal{P}(\Omega)$.

§3.1. The Preservation of Continuity and Condensability.

**Lemma 3.1.** If $A_1 \times B_1, \ldots, A_n \times B_n$ are elements of $\mathcal{E}$, then

$$\mu(\bigvee_{i=1}^n \zeta_1(A_i) \wedge \zeta_2(B_i)) = \sum_{\mathclap{I \subset \{1, \ldots, n\}}}^\sim f_1^*(\bigcup A_i)\mu(\wedge \zeta_2(B_i) - \bigvee \zeta_2(B_i)). \tag{3.8}$$

**Proof:**

$$\mu(\bigvee_{i=1}^n \zeta_1(A_i) \wedge \zeta_2(B_i)) = \sum_{\mathclap{I \subset \{1, \ldots, n\}}}^\sim (-1)^{|I|+1} \mu(\wedge \zeta_1(A_i) \wedge \zeta_2(B_i))_{\mathclap{I \neq \emptyset}}$$

$$= \sum_{\mathclap{I \subset \{1, \ldots, n\}}}^\sim (-1)^{|I|+1} \mu(\wedge \zeta_1(A_i))\mu(\wedge \zeta_2(B_i))_{\mathclap{I \neq \emptyset}}$$

$$= \sum_{\mathclap{I \subset \{1, \ldots, n\}}}^\sim (-1)^{|I|+1}[\sum_{\mathclap{R \subset I}}^\sim (-1)^{|R|+1} f_1^*(\bigcup A_i)]_{\mathclap{I \neq \emptyset}}.$$
\[
\{ \sum_{i \in S} (-1)^{|S|} f^*_2(U \cup B_i) \}_{S \subseteq I, I \neq \emptyset} = \sum_{R, S \subseteq \{1, \ldots, n\} \text{ s.t. } R \cup S \subseteq I} \sum_{i \in R, i \in S} f^*_1(U \cup A_i) f^*_1(U \cup B_i)(-1)^{|R|}(-1)^{|S|} \sum_{I \subseteq \{1, \ldots, n\}} (-1)^{|I|+1}.
\]

But
\[
\sum_{I \subseteq \{1, \ldots, n\}} (-1)^{|I|+1} = \begin{cases} 0 & \text{if } R \cup S \neq \{1, \ldots, n\} \\ (-1)^{n+1} = (-1)^{-n+1} = (-1)^{|R|}(-1)^{|S|}(-1)^{|R \cap S|+1} & \text{if } R \cup S = \{1, \ldots, n\}. \end{cases}
\]

So
\[
\mu\left(\bigvee_{i=1}^n \zeta_1(A_i) \land \zeta_2(B_i)\right) = \sum_{R, S \subseteq \{1, \ldots, n\} \text{ s.t. } R \cup S = \{1, \ldots, n\}} \sum_{I \subseteq \{1, \ldots, n\}} f^*_1(U \cup A_i) f^*_1(U \cup B_i)(-1)^{|R \cap S|+1}.
\]

\[
= \sum_{I \subseteq \{1, \ldots, n\}} \sum_{\emptyset \neq J \subseteq I} (-1)^{|J|+1} f^*_2(U \cup B_i).
\]

Theorem 3.3.

(i) If \(f_1\) and \(f_2\) are continuous, then so is \(f\).

(ii) If \(f_1\) and \(f_2\) are condensable, then so is \(f\).

Proof: (i) Suppose \(f_1\) and \(f_2\) are continuous. To show that \(f\) is continuous, we must show that
\[ \zeta(\bigcup A_i) = \bigvee_{i} \zeta(A_i) \]

whenever \(A_1, A_2, \ldots\) is a sequence of elements of \(G\) and \(\bigcup A_i \in G\).

And since \(G\) consists of finite unions of elements of \(\mathcal{E}\), this will follow if we can show that

\[ \zeta(A \times B) = \bigvee_{i} \zeta(A_i \times B_i), \]

or (see (3.6) above) that

\[ \zeta_1(A) \land \zeta_2(B) = \bigvee_{i} \zeta_1(A_i) \land \zeta_2(B_i) \] (3.9)

whenever \(A \times B, A_1 \times B_1, A_2 \times B_2, \ldots\) are elements of \(\mathcal{E}\) and \(A \times B = \bigcup_{i} A_i \times B_i\).

Notice that the sequence \(A_1 \times B_1, A_2 \times B_2, \ldots\) in (3.9) can be replaced by a disjoint sequence \(C_1 \times D_1, C_2 \times D_2, \ldots\) such that \(A \times B = \bigcup_{i} C_i \times D_i\) and such that each \(C_j \times D_j\) is contained in some \(A_i \times B_i\). And since \(\zeta_1(E) \land \zeta_2(F) \equiv \zeta_1(G) \land \zeta_2(H)\) whenever \(E \times F \subseteq G \times H\), we will have

\[ \zeta_1(A) \land \zeta_2(B) \equiv \bigvee_{i} \zeta_1(A_i) \land \zeta_2(B_i) \equiv \bigvee_{i} \zeta_1(C_i) \land \zeta_2(D_i) . \]

In order to prove (3.9), it suffices, therefore, to show that the right-hand side has measure greater than or equal to that of the left-hand side under the assumption that the \(A_i \times B_i\) are disjoint.

On the assumption that the \(A_i \times B_i\) are disjoint, we have

\[ \chi_{A \times B} = \sum_{i=1}^{\infty} \chi_{A_i \times B_i} \]. And since \(\max \chi_{A_i} = \chi_{\bigcup_{i \in I} A_i}\) for every finite subset \(I\) of natural numbers, we may apply (2.9) to obtain
\[ f_2^*(X_B)X_A(w_i) = f_2^*(X_A(w_i)X_B) = f_2^*(\sum_{i=1}^{\infty} X_{A_i}(w_i)X_{B_i}) = \sup_{n \in \mathbb{N}} f_2^*(\sum_{i=1}^{n} X_{A_i}(w_i)X_{B_i}) \]

\[ = \sup_{n \in \mathbb{N}} \sum_{i=1}^{n} X_{A_i}(w_i) \bigwedge_{i \in I \setminus \emptyset} (\bigwedge_{i \in \mathcal{I}} Z_{B_i}) \bigvee_{i \in \mathcal{I}} Z_{B_i} \]

for all \( w_i \in \Omega_1 \), or

\[ f_2^*(X_B)X_A = \sup_{n \in \mathbb{N}} \sum_{I \subseteq \{1, \ldots, n\}} X_{A_i} \bigwedge_{i \in I \setminus \emptyset} (\bigwedge_{i \in \mathcal{I}} Z_{B_i}) \bigvee_{i \in \mathcal{I}} Z_{B_i} \]

the supremum being over an increasing sequence of positive functions on \( \Omega_1 \). Hence

\[ \mu(\mathcal{L}_1(A) \land \mathcal{L}_2(B)) = \mu(\mathcal{L}_1(A)) \mu(\mathcal{L}_2(B)) \]

\[ = f_1^*(A) f_2^*(B) = f_1^*(X_A) f_2^*(X_B) \]

\[ = f_1^*(f_2^*(X_B)X_A) \]

\[ = \sup_{n \in \mathbb{N}} f_1^*(\sum_{I \subseteq \{1, \ldots, n\}} X_{A_i} \bigwedge_{i \in I \setminus \emptyset} (\bigwedge_{i \in \mathcal{I}} Z_{B_i}) \bigvee_{i \in \mathcal{I}} Z_{B_i}) \]

\[ \leq \sup_{n \in \mathbb{N}} \sum_{I \subseteq \{1, \ldots, n\}} f_1^*(X_{A_i}) \mu(\bigwedge_{i \in \mathcal{I}} Z_{B_i}) \bigvee_{i \in \mathcal{I}} Z_{B_i} \]

\[ = \mu(\bigvee_{i=1}^{\infty} (\mathcal{L}_1(A_i) \land \mathcal{L}_2(B_i))) . \]

(ii) Suppose \( f_1 \) and \( f_2 \) are condensable, and consider an element \( A = \bigcup_{i=1}^{n} A_i \times B_i \) in \( G \). We have
\[ \zeta(A) = \bigvee_{i=1}^{n} \left( \bigvee_{w_1 \in A_i} \zeta_1(\{w_1\}) \right) \land \left( \bigvee_{w_2 \in B_i} \zeta_2(\{w_2\}) \right) \]

\[ = \bigvee_{1 \leq i \leq n} \bigvee_{(w_1, w_2) \in A_i \times B_i} (\zeta_1(\{w_1\}) \land \zeta_2(\{w_2\})) \]

\[ = \bigvee_{(w_1, w_2) \in A} \zeta(\{w_1, w_2\}) . \]

So for every \( \varepsilon > 0 \) there is a finite set \( B \in G \) such that \( f^*(A) - f^*(B) < \varepsilon \). And hence, by Theorem 5.1 of Shafer (1979), \( f \) is condensable. \( \square \)

Let us call the canonical extension of \( f \) to \( P(\Omega_1 \times \Omega_2) \) (which is also the canonical extension of \( f_0 \) to \( P(\Omega_1 \times \Omega_2) \)) the product of \( f_1 \) and \( f_2 \); we shall denote it by \( f_1 \times f_2 \). Notice that if \( f_1 \) and \( f_2 \) are condensable, then \( f_1 \times f_2 \) is also condensable; for since the algebra \( G \) includes the cofinite subsets of \( \Omega \), the canonical extension of the condensable belief function \( f \) to \( P(\Omega_1 \times \Omega_2) \) coincides with the canonical condensable extensions—it is, in fact, the only condensable extension. (It is also the canonical condensable extension of \( f_0 \) to \( P(\Omega_1 \times \Omega_2) \), but not the only condensable extension of \( f_0 \) to \( P(\Omega_1 \times \Omega_2) \).) However, if \( f_1 \) and \( f_2 \) are merely continuous, then the canonical extension \( f_1 \times f_2 \) may fail to coincide with the canonical continuous extension of \( f \) and hence fail to be continuous. (This is clear from Theorem 3.4 below.) So whenever \( f_1 \) and \( f_2 \) are continuous we will call the canonical continuous extension of \( f \) to \( P(\Omega_1 \times \Omega_2) \) the continuous product of \( f_1 \) and \( f_2 \), and denote it by \( f_1 \tilde{\times} f_2 \). Notice that \( f_1 \tilde{\times} f_2 \) can also be described as the canonical continuous extension of \( f_0 \) to \( P(\Omega_1 \times \Omega_2) \).
§3.2. The Relation to Measure Theory.

The product belief function generalizes the usual idea of a product probability measure.

**Theorem 3.4.** Suppose $\mathcal{A}_i$ is an algebra of subsets of $\Omega_i$, $i = 1, 2$. And set

$$\mathcal{A} = \left\{ \bigcup_{i=1}^{n} A_i \times B_i \mid n \geq 1; \ A_i \in \mathcal{A}_1, \ B_i \in \mathcal{A}_2 \right\}.$$  

(i) Suppose $P_i$ is a finitely additive probability measure on $\mathcal{A}_i$, $i = 1, 2$. Let $P_1 \times P_2$ denote the product measure on $\mathcal{A}$, and let $P_1^*$, $P_2^*$ and $(P_1 \times P_2)^*$ denote the respective inner contents (= canonical extensions) for these measures. Then $P_1^* \times P_2^* = (P_1 \times P_2)^*$.

(ii) Suppose $\mathcal{A}_i$ is a $\sigma$-algebra and $P_i$ is a countably additive (= continuous) probability measure on $\mathcal{A}_i$, $i = 1, 2$. Let $P_1 \times P_2$ denote the product measure on $\mathcal{A}$, and let $P_1^*$, $P_2^*$ and $(P_1 \times P_2)^*$ denote the respective inner measures (= canonical continuous extensions) for these measures. Then $P_1^* \times P_2^* = (P_1 \times P_2)^*$.

**Proof:** We will prove (i); the proof of (ii) is similar. We retain the notation of the preceding discussion, with $P_1^*$ and $P_2^*$ in the roles of $f_1$ and $f_2$. In particular, we denote $(P_1^* \times P_2^*)|\mathcal{A}$ by $f$.

First notice that $f, f^*$ and $P_1 \times P_2$ all agree on $\mathcal{A}$. Indeed, if $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$, then they all assign the value $P_1(A) P_2(B)$ to $A \times B$. And if $A_1 \times B_1, \ldots, A_n \times B_n$ is a finite disjoint collection of such rectangles, then the additivity of $P_1$ and $P_2$ implies that the $L_2(A_i) \wedge L_2(B_i)$ are disjoint, whence
\[ P_1 \times P_2 (\bigcup A_i \times B_i) = \sum_i P_1(A_i)P_2(B_i) = \sum_i \mu(\mathcal{F}_1(A_i) \land \mathcal{F}_2(B_i)) = \mu(\bigcup_i (A_i \times B_i)) = f^*(\bigcup_i A_i \times B_i). \]

It follows that \( f^* \) and \( P_1 \times P_2 \) agree on \( \emptyset \); since \( P_1 \times P_2 \) is additive, they therefore both equal \( f \).

Now \( (P_1 \times P_2)^* \) is the canonical extension of \( P_1 \times P_2 \) from \( \emptyset \) to \( \mathcal{P}(\Omega_1 \times \Omega_2) \), while \( P_1^* \times P_2^* \) is the canonical extension of \( f \) from \( \mathcal{G} \), which contains \( \emptyset \), to \( \mathcal{P}(\Omega_1 \times \Omega_2) \). It follows that

\[ (P_1 \times P_2)^* \equiv P_1^* \times P_2^* \quad (3.10) \]

and that they are equal if they are equal on \( \mathcal{G} \). We shall show that they are equal on \( \mathcal{G} \) by showing that their upper probability functions are equal on \( \mathcal{G} \)--i.e., that \( (P_1 \times P_2)^*|\mathcal{G} = f^* \).

By (3.10), \( (P_1 \times P_2)^*|\mathcal{G} \equiv f^* \). So it suffices to choose

\[ A = \bigcup_{i=1}^{n} A_i \times B_i \quad \text{in } \mathcal{G} \quad \text{and } \epsilon > 0 \quad \text{and show that } \quad (P_1 \times P_2)^*(A) - f^*(A) < \epsilon. \]

But

\[ (P_1 \times P_2)^*(A) = \inf\{P_1 \times P_2(B) | A \subset B \in \mathcal{G}\}; \]

so it suffices to exhibit an element \( B \in \mathcal{G} \) such that \( A \subset B \) and \( P_1 \times P_2(B) - f^*(A) < \epsilon \). We choose \( C_1, \ldots, C_n \in G_1 \) and \( D_1, \ldots, D_n \in G_2 \) such that \( C_i \supset A_i, D_i \supset B_i, \ P_1(C_i) - P_1^*(A_i) < \frac{\epsilon}{2n} \) and \( P_2(D_i) - P_2^*(B_i) < \frac{\epsilon}{2n} \), and we set \( B = \bigcup_{i=1}^{n} C_i \times D_i \). Then \( A \subset B \in \mathcal{G} \),

\[ f^*(C_i \times D_i) - f^*(A_i \times B_i) = P_1(C_i)P_2(D_i) - P_1^*(A_i)P_2^*(B_i) \Rightarrow \frac{\epsilon}{n}, \]
and hence, by Choquet (1953, p. 172),

\[ P_1 \times P_2(B) - f^*(A) = f^*(B) - f^*(A) \]

\[ \equiv \Sigma_{i=1}^{n} [f^*(C_i \times D_i) - f^*(A_i \times B_i)] \]

\[ \equiv \epsilon . \]

\[ \Box \]

§ 3.3. Tonelli's Inequality.

When \( f_2 \) is a belief function on \( \Omega_2 \) and \( A \subseteq \Omega_1 \times \Omega_2 \), we denote by \( f^*_2(A) \) the function on \( \Omega_1 \) given by

\[ f^*_2(A)(w_1) = f^*_2(\{w_2 \mid (w_1, w_2) \in A\}) . \]

Theorem 3.5. If \( f_i \) is a belief function on \( P(\Omega_i), \ i = 1, 2 \), then

\[ (f_1 \times f_2)^*(A) \equiv f_1^*(f_2^*(A)) \] (3.11)

for all \( A \subseteq \Omega_1 \times \Omega_2 \). If \( f_1 \) and \( f_2 \) are continuous, then

\[ (f_1 \times f_2)^*(A) \equiv f_1^*(f_2^*(A)) \] (3.12)

for all \( A \subseteq \Omega_1 \times \Omega_2 \).

Proof: First consider an element \( B \in G \). We can write \( B = \bigcup_{i=1}^{n} A_i \times B_i \), where the \( A_i \) are disjoint. In this case

\[ f^*_2(B) = \Sigma_{i=1}^{n} f^*_2(B_i) \times A_i . \]

So by (2.9) and (3.8),
\[ f_1^*(f_2^*(B)) = \sum_{I \subseteq \{1, \ldots, n\}} \left( \max_{i \in I} f_2^*(B_i) \right) \mu(\bigwedge_{i \in I} \mu_1(A_i) - \bigvee_{i \in I} \mu_1(A_i)) \]
\[ \equiv \sum_{I \subseteq \{1, \ldots, n\}} f_2^*(\bigcup_{i \in I} B_i) \mu(\bigwedge_{i \in I} \mu_1(A_i) - \bigvee_{i \in I} \mu_1(A_i)) \]
\[ = (f_1 \times f_2)^*(B) . \]

And we find the value of \((f_1 \times f_2)^*\) for an arbitrary subset \(A\) of \(\Omega_1 \times \Omega_2\) by taking the infima over all \(B \in \mathcal{G}\) such that \(A \subseteq B\):

\[ (f_1 \times f_2)^*(A) = \inf(f_1 \times f_2)^*(B) \equiv \inf f_1^*(f_2^*(B)) \equiv f_1^*(\inf f_2^*(B)) \equiv f_1^*(f_2^*(A)) . \]

So (3.11) holds.

Now suppose \(f_1\) and \(f_2\) are continuous, and let \(\tilde{\mathcal{G}}\) denote the subset of \(\mathcal{P}(\Omega)\) obtained by closing \(\mathcal{G}\) under countable intersections.

An element \(B\) of \(\tilde{\mathcal{G}}\) can always be represented as the union of an increasing sequence \(B_1 \subseteq B_2 \subseteq \ldots\) of elements of \(\mathcal{G}\); and

\[ (f_1 \times f_2)^*(B) = \lim_{i \to \infty} (f_1 \times f_2)^*(B_i) \equiv \lim_{i \to \infty} f_1^*(f_2^*(B_i)) = f_1^*(f_2^*(B)). \]

Since the value of \((f_1 \times f_2)^*\) for an arbitrary subset \(A\) of \(\Omega_1 \times \Omega_2\) is found by taking the infima over all \(B \in \tilde{\mathcal{G}}\) such that \(A \subseteq B\), (3.12) follows.

In the case where \(f_1\) and \(f_2\) are measures and \(A\) is measurable, the Tonelli's Theorem (see, e.g., Bartle, 1966, p. 118) says that (3.12) holds with equality. As the simple example below demonstrates, one cannot expect equality in the case of belief functions.
Example. Set $\Omega_1 = \{a, b\}$ and $\Omega_2 = \{c, d\}$. Let $f_1$ be the vacuous belief function on $\Omega_1$ and let $f_2$ be the probability measure on $\Omega_2$ that assigns probability $\frac{1}{2}$ to each of $c$ and $d$. Let $A = \{(a, c), (b, d)\}$. Then the function $f_2^*(A)$ is identically equal to the constant $\frac{1}{2}$, and hence $f_1^*(f_2^*(A)) = \frac{1}{2}$. But $(f_1 \times f_2)^*(A) = 1$.

§4. Evidential Independence.

The notion of product belief functions can be immediately generalized to the case of belief functions defined on independent subalgebras of an algebra of sets. Such a generalization is of interest primarily because it affords a convenient notation for working with subsets qua propositions, especially when one passes from one algebra to another. In this section we exploit this notational advantage to study the notion of evidential independence.

Recall that two subalgebras $\mathcal{G}_1$ and $\mathcal{G}_2$ of an algebra $\mathcal{G}$ of subsets of a set $\Omega$ are said to be independent if $\emptyset \notin A \in \mathcal{G}_1$ and $\emptyset \notin B \in \mathcal{G}_2$ imply that $A \cap B \notin \emptyset$. (Equivalently: if $A \in \mathcal{G}_1$, $B \in \mathcal{G}_2$, and $A \subseteq B$, then either $A = \emptyset$ or $B = \Omega$.) Let us notice how the construction of the product $f_1 \times f_2$ generalizes to the case where $f_1$ is a belief function on $\mathcal{G}_1$ and $f_2$ is a belief function on $\mathcal{G}_2$. We set

$$\mathcal{E} = \{A \cap B \mid A \in \mathcal{G}_1 : B \in \mathcal{G}_2\} \subseteq \mathcal{G},$$

notice that a non-empty element of $\mathcal{E}$ is uniquely expressible in the form $A \cap B$ with $A \in \mathcal{G}_1$ and $B \in \mathcal{G}_2$, and define $f_0$ on $\mathcal{E}$ by
\[ f_0(A \cap B) = f_1(A) f_2(B). \]

The construction of Theorem 3.1 can then proceed; we find that \( f_0 \) is a belief function and define \( f_1 \times f_2 \) to be its canonical extension to \( \mathcal{G} \). Notice that (3.5), the formula for the allocation \( \rho \) for \( f_1 \times f_2 \), becomes

\[ \rho(A) = \bigvee \{ \rho_1(A_1) \land \rho_2(A_2) \mid A_1 \in \mathcal{G}_1; A_2 \in \mathcal{G}_2; A_1 \cap A_2 \subseteq A \} \quad (4.1) \]

and (3.7) is similarly modified. The product \( f_1 \times f_2 \) is an extension of both \( f_1 \) and \( f_2 \): \( (f_1 \times f_2) \mid \mathcal{G}_1 = f_1 \) and \( (f_1 \times f_2) \mid \mathcal{G}_2 = f_2 \).

Theorem 3.2 also continues to hold, but now the conclusions can be written

\[ f(A \cap B) = f(A)f(B) \quad (4.2) \]

and

\[ f^*(A \cap B) = f^*(A)f^*(B) \quad (4.3) \]

for all \( A \in \mathcal{G}_1 \) and \( B \in \mathcal{G}_2 \), where we have written \( f \) for \( f_1 \times f_2 \).

And in the case where the algebra \( \mathcal{G} \) is generated by \( \mathcal{E} \), we obtain a converse to Theorem 3.2.

**Theorem 3.1.** Suppose \( f \) is a belief function on an algebra \( \mathcal{G} \) of subsets of a set \( \Omega \), suppose \( \mathcal{G}_1 \) and \( \mathcal{G}_2 \) are independent subalgebras of \( \mathcal{G} \), and suppose (4.2) and (4.3) hold for all \( A \in \mathcal{G}_1 \) and \( B \in \mathcal{G}_2 \).

(i) If \( \mathcal{G} \) is the algebra generated by \( \mathcal{G}_1 \cup \mathcal{G}_2 \), then

\[ f = (f \mid \mathcal{G}_1) \times (f \mid \mathcal{G}_2). \]

(ii) If \( f \) is condensable, \( \mathcal{G}_1 \) and \( \mathcal{G}_2 \) are complete, and \( \mathcal{G} \) is the complete algebra generated by \( \mathcal{G}_1 \cup \mathcal{G}_2 \), then \( f = (f \mid \mathcal{G}_1) \times (f \mid \mathcal{G}_2) \).
Proof. (i) Let \( \rho : \mathcal{G} \rightarrow \mathcal{M} \) be an allocation for \( f \). If \( A \in \mathcal{G}_1 \) and \( B \in \mathcal{G}_2 \), then
\[
1-f(A \cup B) = f^*(\overline{A} \cap \overline{B}) = f^*(A)f^*(B)
\]
\[
= (1-f(A))(1-f(B))
\]
\[
= 1-f(A)-f(B)+f(A \cap B),
\]
or
\[
f(A \cup B) = f(A) + f(B)-f(A \cap B),
\]
or
\[
\mu(\rho(A \cup B)) = \mu(\rho(A) \lor \rho(B)).
\]
Hence
\[
\rho(A \cup B) = \rho(A) \lor \rho(B) \tag{4.4}
\]
for all \( A \in \mathcal{G}_1 \) and \( B \in \mathcal{G}_2 \).

Since \( \mathcal{G} \) is generated by \( \mathcal{G}_1 \cup \mathcal{G}_2 \), an arbitrary element \( A \in \mathcal{G} \) can be written in the form
\[
A = (A_1 \cup B_1) \cap \ldots \cap (A_n \cup B_n).
\]
with the \( A_i \) in \( \mathcal{G}_1 \) and the \( B_i \) in \( \mathcal{G}_2 \). Using (4.4), this yields
\[
\rho(A) = [\rho(A_1) \lor \rho(B_1)] \land \ldots \land [\rho(A_n) \lor \rho(B_n)]
\]
\[
= \lor_{I \subseteq \{1, \ldots, n\}} \lor_{i \in I} \land_{i \in I} \rho(A_i \cap B_i).
\]
(Here we use the conventions that \( \land_{i \in \emptyset} = \Omega \) and \( \lor_{i \in \emptyset} = \emptyset \).)

But
\[
\land_{i \in I} A_i \cap \land_{i \in I} B_i \subseteq A
\]
for all \( I \subseteq \{1, \ldots, n\} \). Hence
\[ p(A) = \bigvee \{ p(A_1 \cap A_2) \mid A_1 \in \mathcal{G}_1; A_2 \in \mathcal{G}_2; A_1 \cap A_2 \subseteq A \} \]

for all \( A \in \mathcal{G} \). It is easily verified that the measure of the right-hand side of this equation equals the measure of the right-hand side of (4.1)—i.e., that \( f(A) = (f|_{\mathcal{G}_1}) \times (f|_{\mathcal{G}_2})(A) \).

(ii) Since the algebras are complete, they are isomorphic to power sets. In fact, we can assume, without loss of generality, that \( \mathcal{G} \) is a power set \( \mathcal{P}(\Omega_1 \times \Omega_2) \) and \( \mathcal{G}_1 \) is the subalgebra of \( \mathcal{P}(\Omega_1 \times \Omega_2) \) corresponding to \( \mathcal{P}(\Omega_1) \)—i.e., \( \mathcal{G}_1 = \{ A \times \Omega_2 \mid A \subseteq \Omega_1 \} \) and \( \mathcal{G}_2 = \{ \Omega_1 \times B \mid B \subseteq \Omega_2 \} \). Let \( \mathcal{G}_o \) denote the algebra generated by \( \mathcal{G}_1 \times \mathcal{G}_2 \); then by part (1) of the theorem, \( f|_{\mathcal{G}_o} = ((f|_{\mathcal{G}_1}) \times (f|_{\mathcal{G}_2}))|_{\mathcal{G}_o} \).

But \( \mathcal{G}_o \) includes all cofinite subsets of \( \Omega_1 \times \Omega_2 \) and hence a condensable belief function on \( \mathcal{P}(\Omega_1 \times \Omega_2) \) is uniquely determined by its values on \( \mathcal{G}_o \). It follows that \( f = (f|_{\mathcal{G}_1}) \times (f|_{\mathcal{G}_2}) \). \( \square \)

The independent subalgebras \( \mathcal{G}_1 \) and \( \mathcal{G}_2 \) are called **evidentially independent** with respect to \( f \) when (4.2) and (4.3) hold. As the theorem demonstrates, the name can be interpreted to mean that the evidence affecting \( \mathcal{G}_1 \) is (or at least could be) independent of the evidence affecting \( \mathcal{G}_2 \). Notice that evidential independence reduces to the usual notion of "probabilistic independence" when \( f \) is a probability measure.
§5. Dempster’s Rule of Conditioning.

Suppose we begin with a belief function \( f \) on \( \mathcal{P}(\Omega) \) and then obtain new evidence whose precise and full effect, insofar as it bears on \( \Omega \), is to show that the truth must lie in a certain proper subset \( A \) of \( \Omega \). How should we change \( f \) in order to reflect this new evidence? The idea that each of our degrees of belief represents the measure of a portion of our belief readily leads to an answer to this question.

Recall that if \( \rho : \mathcal{P}(\Omega) \to \mathbb{R} \) is an allocation for \( f \), then the elements of \( \mathcal{M} \) are thought of as portions of our belief or as "probability masses"; \( \rho(B) \) is the total portion of belief committed to \( B \subseteq \Omega \), and its measure \( \mu(\rho(B)) = f(B) \) is our degree of belief in \( B \). When we consider how new knowledge of the truth of a particular subset \( A \) should affect how we commit these portions of belief, our first thought is that \( \rho(B) \), the portion of belief previously committed to \( B \), should now be committed to \( B \cap A \). To put it another way, we should now commit to \( B \) all the belief previously committed to any subset \( C \subseteq \Omega \) such that \( C \cap A \subseteq B \). Since

\[
\bigcup \{ C \subseteq \Omega \mid C \cap A \subseteq B \} = B \cup \overline{A},
\]

this comes down to saying that we should commit to \( B \) all the belief previously committed to \( B \cup \overline{A} \)--i.e., we should commit to \( B \) the probability mass \( \rho(B \cup \overline{A}) \) of measure \( f(B \cap \overline{A}) \).

The difficulty with this suggestion lies, of course, in the presence of the probability mass \( \rho(\overline{A}) \), previously committed to \( \overline{A} \), which we now know to be false. It certainly is not appropriate to commit \( \rho(\overline{A}) \).
to $\overline{A} \cap A = \emptyset$, as the preceding paragraph suggests. It seems more appropriate (and consonant with the usual intuitive explanation of the Bayesian treatment of "conditional probability") to discard this particular probability mass and to renormalize the measure of the remainder so as to bring its total measure back up to unity. (This means multiplying the measure of each remaining probability mass by the inverse of the measure of the total mass remaining—i.e., by $(1-f(\overline{A}))^{-1}$.)

Thus we arrive at the following prescription: we commit to $B$ not the probability mass $\rho(B \cup \overline{A})$ but rather the probability mass $\rho(B \cup \overline{A}) - \rho(\overline{A})$, and our new degree of belief in $B$ is its renormalized measure, namely

$$\frac{f(B \cup \overline{A}) - f(\overline{A})}{1-f(\overline{A})}.$$ 

We call this the conditional degree of belief in $B$ given $A$ and denote it by $f(B|A)$.

**Theorem 5.1.** If $f$ is a belief function on $\mathcal{P}(\Omega)$, $A \subseteq \Omega$ and $f^*(A) > 0$, then the function $f(\cdot|A)$ on $\mathcal{P}(\Omega)$, defined by

$$f(B|A) = \frac{f(B \cup \overline{A}) - f(\overline{A})}{1-f(\overline{A})} \quad (5.1)$$

for all $B \subseteq \Omega$, is also a belief function. If $f$ is continuous, then so is $f(\cdot|A)$; if $f$ is condensable, then so is $f(\cdot|A)$. (Notice that $f^*(A)$ must be greater than zero in order for $f(B|A)$ to be defined.)

The easiest way to see the truth of Theorem 5.1 is to examine the upper probability function for $f(\cdot|A)$. This is the function $f^*(\cdot|A)$
on $P(\Omega)$, given by

$$f^*(B|A) = 1 - f(\overline{B}|A) = 1 - \frac{f(B \cup \overline{A}) - f(A)}{1 - f(A)},$$

or

$$f^*(B|A) = \frac{f^*(B \cap A)}{f^*(A)}, \tag{5.2}$$

and it is evident from this formula that $f^*(\Omega|A) = 1$ and $f^*(\emptyset|A) = 0$, and that $f^*(\cdot|A)$ inherits from $f^*$ the property of being alternating of order $\infty$, as well as properties such as continuity or condensability.

Equation (5.1) is called Dempster's rule of conditioning. It is evident from (5.2) that it reduces, in the special case where $f$ is a probability measure, to the usual Bayesian rule of conditioning.

(cf. pp. 44-45 and p. 67 of Shafer (1976a).) It itself is a special case of Dempster's rule of combination.

After we condition on the subset $A$ of $\Omega$, we may wish to continue to consider $\Omega$ as our set of possibilities, or we may wish to put $A$ in that role. Either attitude is possible after applying the rule of conditioning, for though $f(\cdot|A)$ is defined on $P(\Omega)$, it awards the subset $A$ degree of belief one and hence conveys no more information than its restriction to $P(A)$. We shall denote by $f_A$ the belief function obtained by restricting $f(\cdot|A)$ to $P(A)$, and we shall use the term conditional belief function given $A$, as occasion demands, to refer either to $f(\cdot|A)$ or to $f_A$.

It should also be noted that the rule of conditioning (5.1) can be applied to a belief function defined on an algebra of subsets which is not a power set. All the assertions of 5.1 remain true in this case.
The fact that conditioning is impossible if \( f^*(A) = 0 \) (or \( f(\overline{A}) = 1 \)) would occasion no embarrassment if this relation could be interpreted to mean that \( f \) considers it certain that the truth is in \( \overline{A} \). But as we know from our study of probability theory, this interpretation is not always possible; if \( f \) is a countably additive measure on \( \Omega \) that has no atoms, then \( f^*(\{w\}) = 0 \) for all \( w \in \Omega \), but we can hardly interpret this to mean that \( f \) is certain that none of the elements of \( \Omega \) are true. Fortunately, though, the natural interpretation is possible if \( f \) is condensable.

Indeed, if \( f \) is condensable belief function on \( \mathcal{P}(\Omega) \), then it is obvious that the subset \( C \) of \( \Omega \) given by

\[
C = \cap \{ A \in C \mid f(A) = 1 \}
\]

will satisfy \( f(C) = 1 \) and hence will be non-empty. We call \( C \) the core of \( f \); it is the smallest subset of \( \Omega \) to which \( f \) assigns degree of belief one, and \( \overline{C} \) is the largest subset of \( \Omega \) to which \( f \) assigns upper probability zero. Since \( \overline{C} \) is a proper subset \( \Omega \), it is possible to interpret it as the set of points \( f \) is certain are false; since \( f^*(A) = 0 \) implies \( A \subset \overline{C} \), the failure of the rule of conditioning when this relation holds is then quite natural.

§6. **Dempster's Rule of Combination.**

Dempster's rule of combination is a rule for combining two belief functions \( f_1 \) and \( f_2 \), both defined on the same power set \( \mathcal{P}(\Omega) \), to obtain a new belief function on \( \mathcal{P}(\Omega) \), the orthogonal sum \( f_1 \oplus f_2 \).

This rule was first introduced by Dempster (1966), though special cases
were adduced by Bernoulli (1713) and Lambert (1764). (See Shafer (1978b) Shafer (1976a) studies the rule in detail for the case of finite \( \Omega \) and concludes that it corresponds to the pooling of the evidence underlying \( f_1 \) and \( f_2 \) provided two conditions are met: (i) the sources of evidence must be "independent"--i.e., e, the evidence underlying \( f_1 \) must be entirely distinct from the evidence underlying \( f_2 \)--and (ii) the set of possibilities \( \Omega \) must be sufficiently fine to discern the relevant interaction of the two bodies of evidence.

The general rule of combination can easily be described using the tools we now have at hand: Given \( f_1 \) and \( f_2 \) on \( \mathcal{P}(\Omega) \), we first form the product \( f_1 \times f_2 \) on \( \mathcal{P}(\Omega \times \Omega) \). We then condition \( f_1 \times f_2 \) on the diagonal \( \emptyset \) of \( \Omega \times \Omega \); \( f_1 \Theta f_2 \) is the result, considered as a belief function on \( \mathcal{P}(\Omega) \). (Conditioning \( f_1 \times f_2 \) on \( \emptyset \) results formally in a belief function on \( \mathcal{P}(\emptyset) \). But \( \emptyset \) and \( \Omega \) can be identified in a natural way.) This recipe requires, of course, that \( f_1 \times f_2 \) award the diagonal \( \emptyset \) positive upper probability; if it does not so, then we say that \( f_1 \Theta f_2 \) does not exist.

The belief function \( f_1 \Theta f_2 \) can easily be expressed in terms of the allocations \( \rho_1 \) and \( \rho_2 \) constructed in §3; we have

\[
(f_1 \Theta f_2)(A) = \frac{\mu(\delta(A)) - \mu(\delta(\emptyset))}{1 - \mu(\delta(\emptyset))} ,
\]

where

\[
\delta(A) = \bigvee \{ \rho_1(A_1) \wedge \rho_2(A_2) \mid A_1, A_2 \subset \Omega; A_1 \cap A_2 \subset A \} \quad (6.2)
\]

\( f_1 \Theta f_2 \) exists, of course, if and only if \( \mu(\delta(\emptyset)) < 1 \).
Notice that we present this rule as a definition; we do not attempt to derive it from simpler axioms. Many such axiomatic derivations are not doubt possible; one might be based on the notion of weights of evidence. (See §4.3 of Shafer (1976a).) But the idea that the rule corresponds to the pooling of evidence must find its fundamental justification in examples and in the meaningfulness of the general theory that the rule generates. For the rudiments of such a justification, the reader is again referred to Shafer (1976a).

Since Dempster's rule of combination is a composition of the rule of conditioning and the rule for forming products, it is not surprising that it reduces to those rules in special cases.

**Theorem 6.1.** (i) Suppose \( f_1 \) is a belief function \( P(\Omega) \), \( A \subseteq \Omega \), and \( f_2 \) is the belief function on \( P(\Omega) \) given by

\[
f_2(B) = \begin{cases} 
1 & \text{if } A \subseteq B \\
0 & \text{if } A \not\subseteq B
\end{cases}
\]

for all \( B \subseteq \Omega \). Then \( f_1 \odot f_2 \) exists if and only if \( f_1^*(A) > 0 \). If \( f_1 \odot f_2 \) exists, then \( f_1 \odot f_2 = f(\cdot | A) \).

(ii) Suppose \( f_1 \) and \( f_2 \) are belief functions on \( P(\Omega_1 \times \Omega_2) \), and \( f_1 \) is discerned by \( P(\Omega_1) \), regarded as a subalgebra of \( P(\Omega_1 \times \Omega_2) \). Then \( f_1 \odot f_2 \) exists and is equal to \( (f_1|P(\Omega_1)) \times (f_2|P(\Omega_2)) \).

**Proof:** (i) In this case \( \rho_2 \) is given by

\[
\rho_2(B) = \begin{cases} 
\lor & \text{if } A \subseteq B \\
\land & \text{if } A \not\subseteq B
\end{cases}
\]
Hence (6.2) yields, for any $B \subset \Omega$,

$$
\delta(B) = \bigvee \{ \rho_1(A_1) \mid A_1, A_2 \subset \Omega; A_1 \cap A_2 \subset B; A \subset A_2 \}
= \rho_1(B \cup \overline{A}).
$$

And hence (6.1) yields

$$
(f_1 \otimes f_2)(B) = \frac{f_1(B \cup \overline{A}) - f_1(\overline{A})}{1 - f_1(\overline{A})} = f_1(B \mid A).
$$

(ii) Recall the method of §3 for constructing allocations $\rho_1, \rho_2$ and $\rho$ for $f_1 \mid \mathcal{P}(\Omega_2)$, $f_2 \mid \mathcal{P}(\Omega_2)$, and $(f_1 \mid \mathcal{P}(\Omega_1)) \times (f_2 \mid \mathcal{P}(\Omega_2))$: For $i = 1, 2$, let $(\mathcal{M}_i, \mu_i)$ be a probability algebra and let $r_i : \mathcal{P}(\Omega_i) \rightarrow \mathcal{M}_i$ be an initial allocation for $f_i \mid \mathcal{P}(\Omega_i)$. Construct $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$ and $\mu = \mu_1 \times \mu_2$ as in the proof of Theorem 3.1, denote by $h_i$ the canonical homomorphism of $\mathcal{M}_i$ into $\mathcal{M}$ and set $\rho_i = h_i \circ r_i$. And define $\rho : \mathcal{P}(\Omega_1 \times \Omega_2) \rightarrow \mathcal{M}$ by (3.5).

Now let $\overline{r}_i$ and $\overline{\rho}_i$ be the canonical extensions to $\mathcal{P}(\Omega_x \Omega_2)$ of $r_i$ and $\rho_i$ respectively, $i = 1, 2$. Notice that $\overline{\rho}_i = h_i \circ \overline{r}_i$. Since $f_i$ is discerned by $\mathcal{P}(\Omega_i)$, both $\overline{r}_i$ and $\overline{\rho}_i$ are allocations for $f_i$. And since $\overline{\rho}_i = h_i \circ \overline{r}_i$, the $\overline{\rho}_i$ can be used to define an allocation $\overline{\rho}$ for $f_1 \times f_2$:

$$
\overline{\rho}(A) = \bigvee \{ \overline{\rho}_1(A_1) \wedge \overline{\rho}_2(A_2) \mid A_1, A_2 \subset \Omega_1 \times \Omega_2; A_1 \times A_2 \subset A \}
$$

for all $A \subset (\Omega_1 \times \Omega_2) \times (\Omega_1 \times \Omega_2)$.

Applying (6.1) and (6.2) to the allocation $\overline{\rho}$, we find that $f_1 \otimes f_2$ is given by (6.1), where

$$
\delta(A) = \bigvee \{ \overline{\rho}_1(A_1) \wedge \overline{\rho}_2(A_2) \mid A_1, A_2 \subset \Omega_1 \times \Omega_2; A_1 \cap A_2 \subset A \}.
$$
But this formula reduces to

\[ \delta(A) = \bigvee \{ (\rho_1(B_1) \mid B_1 \subset \Omega_1 ; B_1 \times \Omega_2 \subset A_1) \} \]
\[ \land \{ \bigvee \{ \rho_2(B_2) \mid B_2 \subset \Omega_2 ; \Omega_1 \times B_2 \subset A_1 \} \mid A_1, A_2 \subset \Omega_1 \times \Omega_2 ; A_1 \cap A_2 \subset A \} \]
\[ = \bigvee \{ \bigvee \{ \rho_1(B_1) \land \rho_2(B_2) \mid B_1 \subset \Omega_1 ; B_2 \subset \Omega_2 ; B_1 \times \Omega_2 \subset A_1 ; A_2 \subset \Omega_1 \times \Omega_2 ; A_1 \cap A_2 \subset A \} \} \]
\[ = \bigvee \{ \rho_1(B_1) \land \rho_2(B_2) \mid B_1 \subset \Omega_1 ; B_2 \subset \Omega_2 , B_1 \times B_2 \subset A \} \]
\[ = \rho(A) . \]

Hence (6.1) yields

\[ (f_1 \otimes f_2)(A) = \mu(\rho(A)) = [(f_1|\mathcal{P}(\Omega_1)) \times (f_2|\mathcal{P}(\Omega_2))](A) . \]

When does \( f_1 \otimes f_2 \) exist? It obviously never exists when \( f_1 \) and \( f_2 \) are inner measures for measures that have no atoms, and this casts considerable doubt on the meaningfulness of the rule of combination for belief functions that are merely continuous. But matters are quite different when at least one of the belief functions being combined is condensible and the other is continuous.

**Theorem 6.2.** Suppose \( f_1 \) and \( f_2 \) are belief functions on \( \mathcal{P}(\Omega) \), \( f_1 \) is continuous and \( f_2 \) is condensible. And suppose that \( f_1 \) and \( f_2 \) do not flatly contradict each other—i.e., there does not exist any subset \( A \) of \( \Omega \) such that \( f_1(A) = f_2(\overline{A}) = 1 \). Then \( f_1 \otimes f_2 \) exists.
Proof: Let \( C \) denote the core of \( f_2 \). Then \( f_1^*(C) > 0 \) by hypothesis. And

\[
f_2^*(\emptyset)(w) = f_2^*(\{w\}) > 0
\]

for all \( w \in C \). Since \( f_1 \) is continuous, it follows that the functional \( f_1^* \) must assign a positive value to the function \( f_2^*(\emptyset) \). So Tonelli's inequality yields

\[
(f_1 \times f_2)^*(\emptyset) \leq f_1^*(f_2^*(\emptyset)) > 0.
\]

Notice that if both \( f_1 \) and \( f_2 \) are condensable, then the statement that the two do not flatly contradict each other is equivalent to the statement that their cores intersect.

Theorem 6.2 has an intuitive interpretation. Because of their connection with "weights of evidence" (see the discussions of the commonality function in Shafer (1976a and b)), condensable belief functions seem appropriate for the representation of empirical evidence. Belief functions that are merely continuous, on the other hand, are typified by the continuous probability measures used to represent chances or "objective probabilities"; such probability measures represent theoretical knowledge, knowledge which can be tested empirically but which does not pretend to be merely a representation of empirical evidence. Thus the theorem corresponds to saying that empirical and theoretical knowledge can be combined. And we need not be concerned that two continuous belief functions may fail to be combinable--this corresponds to saying that two rival theoretical systems may fail to be combinable.
Since conditioning and the formation of products preserves condensability, $f_1 \otimes f_2$ will be condensable if $f_1$ and $f_2$ are. But it is evident from part (ii) of Theorem 6.1 that $f_1 \otimes f_2$ need not be continuous just because $f_1$ and $f_2$ are. One might conjecture that $f_1 \times f_2$ and hence $f_1 \otimes f_2$ will be continuous if $f_1$ is continuous and $f_2$ is condensable, but it is an open question whether this is so. We can, of course, construct a continuous orthogonal sum of merely continuous belief functions by replacing $f_1 \times f_2$ by $f_1 \bar{\times} f_2$ in the recipe for constructing $f_1 \otimes f_2$. This replacement will not affect the truth or the proof of Theorem 6.2, for Tonelli's inequality will also hold for $f_1 \bar{\times} f_2$.

§ 7. **Cognitive Independence.**

The rule of combination easily generalizes to the case of belief functions on an algebra of subsets that is not a power set. In this section we use this generalization to study the notion of cognitive independence.

There are several intuitive approaches to generalizing the rule of combination to the case of two belief functions $f_1$ and $f_2$ defined on a proper subset $\mathcal{G}$ of a power set $\mathcal{P}(\Omega)$. Here we use the simplest: we canonically extend $f_1$ and $f_2$ to $\mathcal{P}(\Omega)$, combine, and then restrict to $\mathcal{G}$ again. This results in obvious modifications in the formulae in § 6. Equations (6.1) and (6.2) become

$$
(f_1 \otimes f_2)(A) = \frac{\mu(\delta(A)) - \mu(\delta(\emptyset))}{1 - \mu(\delta(\emptyset))}
$$

(7.1)
where
\[
\delta(A) = V\{\rho_1(A_1) \wedge \rho_2(A_2) \mid A_1, A_2 \in G; A_1 \cap A_2 \subseteq A\}. \tag{7.2}
\]
and Theorem 6.1 (i) remains valid.

(An obvious question arises: If \( f_1 \) and \( f_2 \) are belief functions on \( P(\Omega) \) and \( G \) is a proper subalgebra of \( P(\Omega) \), then will \( (f_1|G) \Theta (f_2|G) = (f_1 \Theta f_2)|G \)? The obvious answer: Generally not, unless \( f_1 \) and \( f_2 \) are discerned by \( G \). The rule of combination is sensible only if the subalgebra \( G \), and more fundamentally the set of possibilities \( \Omega \), is sufficiently fine to discern the relevant interaction of the two bodies of evidence. For an extensive discussion, see Chapter 8 of Shafer (1976a).)

Two independent subalgebras \( G_1 \) and \( G_2 \) of an algebra of subsets \( G \) are said to cognitively independent with respect to a belief function \( f \) on \( G \) if
\[
f^*(A \cap B) = f^*(A)f^*(B) \tag{7.3}
\]
for all \( A \in G_1 \) and \( B \in G_2 \). The intuitive content of this name is explained by the following theorem.

Theorem 7.1. Suppose \( G_1 \) and \( G_2 \) are independent subalgebras of \( G \) and suppose \( f \) is a belief function on \( G \). Then the following assertions are equivalent:

1. \( G_1 \) and \( G_2 \) are cognitively independent with respect to \( f \).
2. \( f(A \mid B) = f(A) \) whenever \( A \in G_1 \), \( B \in G_2 \), and \( f^*(B) > 0 \).
3. If \( f_2 \) is a belief function on \( G \) that is discerned by \( G_2 \), and \( f \Theta f_2 \) exists, then \( (f \Theta f_2)|G_1 = f|G_1 \).
Proof. The equivalence of (1) and (2) is obvious, especially when the equation in (2) is rewritten as \( f^*(A|B) = f^*(A) \).

Furthermore, (2) is obviously a special case of (3). For if we choose \( B \in G_2 \) such that \( f^*(B) > 0 \) and consider the belief function \( f_2 \) given by

\[
f_2(C) = \begin{cases} 
1 & \text{if } B \subseteq C \\
0 & \text{if } B \not\subseteq C,
\end{cases}
\]

then Theorem 6.1 tells us that \( f \, \Theta \, f_2 = f(\cdot | B) \). And hence the conclusion of (3) is that \( f(\cdot | B)|G_1 = f|G_1 \).

To complete the proof, let us assume that (7.3) holds and deduce (3). First note that (7.3) is equivalent to

\[
f(A \cup B) - f(A) = (1 - f(A))f(B)
\]

for all \( A \in G_1 \) and \( B \in G_2 \). Let \( f_2 \) be a belief function on \( G \) that is discerned by \( G_2 \), fix \( A \in G_1 \), and calculate \( (f \, \Theta \, f_2)(A) \) using (7.1) and (7.2). In the present case

\[
\rho_2(A_2) = \bigvee \{ \rho_2(B) \mid B \in G_2 ; B \subseteq A_2 \},
\]

whence

\[
\delta(A) = \bigvee \{ \rho_1(A \cup \overline{B}) \land \rho_2(B) \mid B \in G_2 \}
\]

\[
= \rho_1(A) \lor \bigvee_{B \in G_2} [\rho(A \cup \overline{B}) - \rho_1(A)] \land \rho_2(B),
\]

and

\[
\mu(\delta(A)) = f(A) + \sup_{B_1 \cdots B_n \in G_2} \mu(\bigvee_{i=1}^n [\rho_1(A \cup \overline{B_i}) - \rho_1(A)] \land \rho_2(B)).
\]

Using (4.2) and (7.4), we find that
\[
\mu(\bigvee_{i=1}^{n} [\rho_1(A \cup \overline{B_i}) - \rho_1(A)] \land \rho_2(B_i)) = \sum_{I \subseteq \{1, \ldots, n\}} \prod_{i \in I} (-1)^{|I|} [f(A \cup \overline{B_i}) - f(A)] f(\cap B_i) \prod_{i \in I} \nu_i \in \mathcal{I} \\
= \sum_{I \subseteq \{1, \ldots, n\}} (-1)^{|I|} [f(A \cup \overline{B_i}) - f(A)] f(\cap B_i) \prod_{i \in I} \nu_i \\
= (1-f(A)) \mu(\bigvee_{i=1}^{n} \rho_1(\overline{B_i}) \land \rho(B_i)) .
\]

Hence

\[\mu(\delta(A)) = f(A) + (1-f(A))\mu(\delta(\emptyset)) .\]

Substituting this in (7.1), we obtain \((f \circ f_2)(A) = f(A)\). Hence (3) holds.

In words: \(G_1\) and \(G_2\) are cognitively independent if new evidence that bears only on \(G_2\) cannot change one's degree of belief about \(G_1\).

Cognitive independence is a weaker notion than evidential independence, for it requires only the second of the two relations (5.1) and (5.2) required by evidential independence. That two subalgebras can in fact be cognitively independent without being evidentially independent is demonstrated by an example in §7.5 of Shafer (1976a).
§8. References.


Lambert, Johann Heinrich (1764). Neues Organon.


