CHAPTER 4. THE MATHEMATICAL REPRESENTATION OF OUR PROBABILITY

Now that we have a better technical grasp of the theory of Boolean algebras, we can improve the mathematical representation of our intuitive "probability masses." In this chapter, that representation is improved and developed.

1. Probability Algebras

In section 1 of Chapter 2, I gave the following definition of a measure on a Boolean algebra:

Definition. If $\mathcal{M}$ is a Boolean algebra, then a function $\mu : \mathcal{M} \to [0, 1]$ is a measure if

(1) $\mu(\Lambda_{\mathcal{M}}) = 0$,

(2) $\mu(V_{\mathcal{M}}) = 1$,

and (3) $\mu(M_1) + \mu(M_2) = \mu(M_1 \vee M_2)$ whenever $M_1, M_2 \in \mathcal{M}$ and $M_1 \wedge M_2 = \Lambda_{\mathcal{M}}$.

I then declared that any Boolean algebra $\mathcal{M}$ with an accompanying measure μ could be called a measure algebra -- the intuitive idea being that the elements of $\mathcal{M}$ could be regarded as probability masses. But as I later observed, there are properties that our "probability masses" ought ideally to have that are not imposed by this definition. At the end of Chapter 2, I listed three such properties: positivity, completeness and complete additivity. Now that we have a stronger technical grasp of the theory of Boolean algebras, we can describe these properties more precisely.

Definition. A measure algebra $(\mathcal{M}, \mu)$ is called a probability algebra if

(1) $(\mathcal{M}, \mu)$ is positive: If $M \in \mathcal{M}$ and $M \neq \Lambda_{\mathcal{M}}$, then $\mu(M) > 0$;

(2) $\mathcal{M}$ is complete;

(3) $(\mathcal{M}, \mu)$ is completely additive: If $\mathcal{C} \subset \mathcal{M}$ and the elements of $\mathcal{C}$ are pairwise disjoint, then $\sum_{M \in \mathcal{C}} \mu(M) = \mu(\vee \mathcal{C})$.

The conditions listed in this definition add up to a rather strong package, and the reader might well question whether there even exist any probability algebras. As it turns out, though, there are quite a few of them. In fact, for every measure algebra $(\mathcal{M}, \mu)$, there exists a probability algebra $(\mathcal{N}, \nu)$ and a Boolean homomorphism $h: \mathcal{M} \to \mathcal{N}$ such that $\mu = \nu \circ h$. This fact will be proven in the next section.

The condition of complete additivity may require some explanation. The symbol $\sum_{M \in \mathcal{C}} \mu(M)$ ostensibly requires the addition of a number of non-negative quantities that may be infinite and perhaps even uncountably infinite. But the sum of an uncountable number of positive quantities does not really exist, or at any rate must be considered infinite, while $\sum_{M \in \mathcal{C}} \mu(M) = \mu(\vee \mathcal{C})$ is supposed to be finite. Hence the condition of complete additivity requires in particular that at most a countable number of the elements of $\mathcal{C}$ can have non-zero measure. If $(\mathcal{M}, \mu)$ is also positive, then this means that only a countable number of the elements of $\mathcal{M}$ can be non-zero. Hence we may conclude that any collection of disjoint non-zero elements in a probability algebra must be countable.

These considerations make the following theorem less surprising than it seems at first:

Theorem. Suppose $(\mathcal{M}, \mu)$ is a measure algebra and satisfies the following conditions:

(i) $\mathcal{M}$ is σ-complete.

(ii) $\mathcal{M}$ is positive.

(iii) $(\mathcal{M}, \mu)$ is countably additive: If $\mathcal{C} \subset \mathcal{M}$, $\mathcal{C}$ is countable and the elements of $\mathcal{C}$ are pairwise disjoint, then
$$\sum_{M \in \mathcal{C}} \mu(M) = \mu(\vee \mathcal{C}).$$

Then $(\mathcal{M}, \mu)$ is a probability algebra.

Proof: It follows from the (finite) additivity of μ that $\mathcal{M}$ cannot contain, for any positive integer n, as many as n elements of measure greater than $1/n$. Hence any disjoint set of elements of $\mathcal{M}$ must be countable.

Let $\mathcal{C}$ be any non-empty subset of $\mathcal{M}$. Then it follows from the second theorem of section 5 of Chapter 3 that there exists a disjoint subset $\mathcal{D}$ of $\mathcal{M}$ with exactly the same upper bounds as $\mathcal{C}$. Since $\mathcal{D}$ is disjoint, it must be countable; and since $\mathcal{M}$ is σ-complete, $\mathcal{D}$ must have a least upper bound or join. The same element will also be the join of $\mathcal{C}$. Hence any non-empty subset of $\mathcal{M}$ has a join. The existence of meets follows; $\mathcal{M}$ is complete. And complete additivity follows from countable additivity. ▨

The fact that any collection of disjoint non-zero elements in a probability algebra must be countable also gives the following interesting result:

Theorem: Suppose $\mathcal{M}$ is a probability algebra, $M \in \mathcal{M}$, $\{M_\gamma\}_{\gamma \in \Gamma}$ is a collection of elements of $\mathcal{M}$ and $M = \vee M_\gamma$. Then there exists a disjoint sequence $M_1, M_2, \ldots$ of elements of $\mathcal{M}$ such that (i) $M = \vee M_i$ and (ii) for each i there exists $\gamma \in \Gamma$ such that $M_i \leq M_\gamma$.

Proof: By the second theorem of section 5 of Chapter 3, there exists a disjoint subset $\mathcal{D}$ of $\mathcal{M}$ with the same set of upper bounds as $\{M_\gamma\}_{\gamma \in \Gamma}$, and such that for each $D \in \mathcal{D}$ there exists $\gamma \in \Gamma$ with $D \leq M_\gamma$. But since $\mathcal{D}$ is disjoint, it can have at most a countably infinite number of non-zero elements. Denoting these by $M_1, M_2, \ldots$ yields the theorem. ▨

A probability algebra also has very strong properties of the type that are often called continuity properties. For a start, the measures of a monotone sequence of elements of the probability algebra will converge to the measure of the limit of the monotone sequence, as shown in the following theorem.

Theorem. Suppose $(\mathcal{M}, \mu)$ is a probability algebra. Then for any monotonically increasing sequence $M_1 \leq M_2 \leq \ldots$ in $\mathcal{M}$,

$$\mu(\vee M_i) = \sup_i \mu(M_i).$$

And for any monotonically decreasing sequence $M_1 \geq M_2 \geq \ldots$ in $\mathcal{M}$,

$$\mu(\wedge M_i) = \inf_i \mu(M_i).$$

Proof: First suppose $M_1 \leq M_2 \leq \ldots$ is an increasing sequence, and set $M_o = \Lambda$. Then it is easily verified from the definition of join that

$$\bigvee_{i=1}^{\infty} M_i = \bigvee_{i=1}^{\infty} (M_i - M_{i-1}).$$

But the elements in the join on the right-hand side are disjoint; hence

$$\mu(\bigvee_{i=1}^{\infty} M_1) = \mu(\bigvee_{i=1}^{\infty} [M_i - M_{i-1}]) = \sum_{i=1}^{\infty} \mu(M_i - M_{i-1})$$

$$= \sup_n \sum_{i=1}^{n} \mu(M_i - M_{i-1}) = \sup_n \mu(\bigvee_{i=1}^{n} (M_i - M_{i-1}))$$

$$= \sup_n \mu(M_n).$$

In the case where $M_1 \geq M_2 \geq \ldots$ is a decreasing sequence, $\overline{M_1} \leq \overline{M_2} \leq \ldots$ an increasing sequence, and $\wedge M_i = \overline{\vee \overline{M_i}}$.

Hence,

$$\mu(\wedge M_i) = \mu(\overline{\vee \overline{M_i}}) = 1 - \mu(\vee \overline{M_i}) = 1 - \sup_i \mu(\overline{M_i})$$

$$= 1 - \sup_i (1 - \mu(M_i)) = \inf_i \mu(M_i).$$ ▨

The proof just given uses the property of additivity only for countable subsets of $\mathcal{M}$. Using the full force of the property of complete additivity, we can prove a rather stronger statement, the formulation of which requires the notion of a net.

A non-empty subset $\mathcal{B}$ of a Boolean algebra $\mathcal{Q}$ is called a downward net in $\mathcal{Q}$ if for every pair of elements A, B ϵ $\mathcal{B}$ there exists an element C ϵ $\mathcal{B}$ such that $C \leq A \wedge B$. A non-empty subset $\mathcal{B}$ of a Boolean algebra is called an upward net in $\mathcal{Q}$ if for every pair of elements A, B ϵ $\mathcal{B}$ there exists an element $C \in \mathcal{B}$ such that $A \vee B \leq C$. Notice that a filter is a downward net and that an ideal is an upward net.

<u>Theorem</u>. Suppose $(\mathcal{M}, \mu)$ is a probability algebra. Then for any downward net $\mathcal{B} \subset \mathcal{M}$,

$$\mu(\wedge \mathcal{B}) = \inf_{B \in \mathcal{B}} \mu(B).$$

And for any upward net $\mathcal{B} \subset \mathcal{M}$,

$$\mu(\vee \mathcal{B}) = \sup_{B \in \mathcal{B}} \mu(B).$$

<u>Proof</u>: Consider first the case of a downward net $\mathcal{B}$. Since $\wedge\mathcal{B} \leq B$ for all $B \in \mathcal{B}$, $\mu(\wedge \mathcal{B}) \leq \inf_{B \in \mathcal{B}} \mu(B)$. Choose a decreasing sequence $B_1 \geq B_2 \geq B_3, \ldots$ in $\mathcal{B}$ such that $\inf_i \mu(B_i) = \inf_{B \in \mathcal{B}} \mu(B)$. Then by the preceding theorem, $\mu(\wedge_i B_i) = \inf_i \mu(B_i) = \inf_{B \in \mathcal{B}} \mu(B)$. Now suppose $\mu(\wedge \mathcal{B}) < \inf_{B \in \mathcal{B}} \mu(B)$. Then $\wedge \mathcal{B}$ is a proper subelement of $\wedge B_i$. This implies the existence of some element $M_1 \in \mathcal{B}$ such that $\wedge B_i$ is not a subelement of M_1, or $\wedge B_i - M_1 \neq \Lambda$. Denote $\mu(\wedge B_i - M_1) = \epsilon > 0$. We can choose an integer K so that $\mu(B_K - \wedge B_i) = \mu(B_K) - \mu(\wedge B_i) < \epsilon/2$, and if we then choose $M_2 \in \mathcal{B}$ so that $M_2 \leq B_K \wedge M_1$, we will have

$$\mu(\wedge B_i - M_2) \geq \epsilon$$

and $$\mu(M_2 - \wedge B_i) < \epsilon/2.$$

This implies that $\mu(\wedge B_i) > \mu(M_2)$, contradicting the assumption that $\mu(\wedge B_i) = \inf_{B \in \mathcal{B}} \mu(B)$. Hence $\mu(\wedge B) = \inf_{B \in \mathcal{B}} \mu(B)$. ▨

If $(\mathcal{M}, \mu)$ is a probability algebra and $\mathcal{N}$ is a complete subalgebra of $\mathcal{M}$, then $(\mathcal{N}, \mu|\mathcal{N})$ will be a probability algebra. We can describe this situation by saying that $(\mathcal{N}, \mu|\mathcal{N})$ is embedded in $(\mathcal{M}, \mu)$. More generally, if $(\mathcal{M}, \mu)$ and $(\mathcal{N}, \nu)$ are probability algebras, then an isomorphism θ of $\mathcal{M}$ into $\mathcal{N}$ is called an embedding or isomorphism of $(\mathcal{M}, \mu)$ into $(\mathcal{N}, \nu)$ if $\mu = \nu \circ \theta$. And of course if θ is also onto, then it is called an isomorphism between the two probability algebras, and they are said to be isomorphic.

2. Constructing Probability Algebras

In this section, I will show that for any measure algebra $(\mathcal{M}, \mu)$ there exists a probability algebra $(\mathcal{N}, \nu)$ and a Boolean homomorphism $h: \mathcal{M} \to \mathcal{N}$ such that $\mu = \nu \circ h$. One important tool in this demonstration is Carathéodory's Extension Theorem, a standard theorem in measure theory that I will state and use without proof.

Carathéodory's Extension Theorem. Suppose $\mathcal{F}$ is a field of subsets of a set $\mathcal{S}$ and $\delta: \mathcal{F} \to [0,1]$ satisfies

(1) $\delta(\phi) = 0$

(2) $\delta(\mathcal{S}) = 1$

(3) $\delta(S_1) + \delta(S_2) = \delta(S_1 \cup S_2)$ whenever $S_1, S_2 \in \mathcal{F}$ and $S_1 \cap S_2 = \phi$.

(4) If $S_1 \supset S_2 \supset \ldots$ and $\bigcap_{i=1}^{\infty} S_i = \phi$, then $\lim_{i \to \infty} \delta(S_i) = 0$.

Let $\mathcal{F}^*$ be the σ- field of subsets of $\mathcal{S}$ generated by $\mathcal{F}$. Then there exists an extension δ^* of δ to $\mathcal{F}^*$ such that

(5) $\sum_{i=1}^{\infty} \delta^*(S_i) = \delta^*(\bigcup_{i=1}^{\infty} S_i)$ for all disjoint sequences $S_1, S_2, \ldots$ of elements of $\mathcal{F}^*$.

Lemma 1. Suppose $(\mathcal{M}, \mu)$ is a measure algebra. Then there exists a measure algebra $(\mathfrak{F}^*, \delta^*)$ that is σ - complete and countably additive and a Boolean homomorphism $f: \mathcal{M} \to \mathfrak{F}^*$ such that $\mu = \delta^* \circ f$.

Proof: Let $f_o: \mathcal{M} \to \mathfrak{F}$ be the isomorphism established by the Stone Representation Theorem; $\mathfrak{F}$ being a field of subsets of the set $\mathcal{S}$ of all ultrafilters in $\mathcal{M}$, with $f_o(M) = \{F \mid F \in \mathcal{S} \text{ and } M \in F\}$. Set $\delta = \mu \circ f_o^{-1}$.

Then $\mathfrak{F}$ and δ obviously satisfy (1), (2) and (3) in the hypothesis of Caratheodory's Extension Theorem. In fact, it also satisfies (4). To see this, let $S_1 \supset S_2 \supset \ldots$ be a decreasing sequence in with $\bigcap S_i = \phi$, and set $M_i = f^{-1}(S_i)$. Then $M_1 \geq M_2 \geq \ldots$, and

$$\phi = \bigcap S_i = \bigcap \{F \mid F \in \mathcal{S} \text{ and } M_i \in F\}$$
$$= \{F \mid F \in \mathcal{S} \text{ and } M_i \in F \text{ for all } i\}.$$

Now set $F_o = \{M \mid M_i \leq M \text{ for some } i\}$. It is easily seen that F_o is a filter. Furthermore, F_o is improper. For if it were proper, it would be contained in an ultrafilter F_1; F_1 would then contain all the M_i and hence would be in S_i, contradicting the assumption that $\bigcap S_i = \phi$. So F_o is improper and thus contains $\Lambda_{\mathcal{M}}$. But this implies that $M_i = \Lambda_{\mathcal{M}}$ for some i and hence for all $j \geq i$. Thus $\lim_{i \to \infty} \delta(S_i) = \lim_{i \to \infty} \mu(M_i) = 0$.

So by Carathéodory's Extension Theorem, δ can be extended to a countably additive measure δ^* on the σ - field $\mathfrak{F}^*$ generated by $\mathfrak{F}$. Evidently, $(\mathfrak{F}^*, \delta^*)$ is a σ - complete and countably additive probability algebra. If we denote by i the identity mapping from $\mathfrak{F}$ into $\mathfrak{F}^*$ then $f = i \circ f_o$ is a Boolean homomorphism of $\mathcal{M}$ into $\mathfrak{F}^*$. Furthermore, $\mu = \delta \circ f_o = \delta^* \circ i \circ f_o = \delta^* \circ f$. ▨

Lemma 2. Suppose $(\mathcal{F}^*, \delta^*)$ is a σ-complete and countably additive measure algebra. Then there exists a probability algebra $(\mathcal{N}, \nu)$ and a Boolean homomorphism $g: \mathcal{F}^* \to \mathcal{N}$ such that $\delta^* = \nu \circ g$.

Proof: Consider the set $I = \{M \mid M \in \mathcal{F}^*, \mu(M) = 0\}$. It is easily shown that I is a proper ideal in $\mathcal{F}^*$. Hence one may construct the quotient $\mathcal{N} = \mathcal{F}^*/I$ and the Boolean homomorphism $g: \mathcal{F}^* \to \mathcal{N}$: $M \rightsquigarrow \{N \mid N \in \mathcal{F}^*, N \Delta M \in I\}$ as in section 8 of Chapter 3. Recall that each element of $\mathcal{N}$ is an equivalence class of elements of $\mathcal{F}^*$. If M and N are both in the equivalence class $E \in \mathcal{N}$, then $N \Delta M \in I$, whence $\delta^*(N \Delta M) = 0$ and $\delta^*(N) = \delta^*(M)$. Hence one may define a function $\nu: \mathcal{N} \to [0, 1]$ by setting $\nu(E) = \delta^*(M)$ for any $M \in E$. Evidently, $\nu = \delta^* \circ g$.

Since $\Lambda_{\mathcal{F}^*}$ is in the equivalence class $\Lambda_{\mathcal{N}}$ and $V_{\mathcal{F}^*}$ is in the equivalence class $V_{\mathcal{N}}$, $\nu(\Lambda_{\mathcal{N}}) = 0$ and $\nu(V_{\mathcal{N}}) = 1$. And if $E_1, E_2 \in \mathcal{N}$ with $E_1 \wedge E_2 = \Lambda_{\mathcal{N}}$, then choosing $M \in E_1$ and $N \in E_2$ gives $g(M \wedge N) = g(M) \wedge g(N) = E_1 \wedge E_2 = \Lambda_{\mathcal{N}} = I$, whence $\delta^*(M \wedge N) = 0$. Hence $\nu(E_1) + \nu(E_2) = \delta^*(M) + \delta^*(N) = \delta^*(M \vee N) = \nu(g(M \vee N)) = \nu(g(M) \vee g(N)) = \nu(E_1 \vee E_2)$. So $(\mathcal{N}, \nu)$ is a probability algebra.

Furthermore, $(\mathcal{N}, \nu)$ is positive. For if $\nu(E) = 0$, then choosing $M \in E$ gives $\delta^*(M) = 0$, whence $M \in I$ and $E = I = \Lambda_{\mathcal{N}}$.

Now I is a σ-ideal. In order to prove this, take any countable collection $A_1, A_2, \ldots$ of elements of I and set $B_i = A_i - \bigvee_{j<i} A_j$, and note that $\vee A_i = \vee B_i$ and $\delta^*(\vee B_i) = \Sigma\, \delta^*(B_i) = 0$. Since I is a σ-ideal, the quotient $\mathcal{N}$ is σ-complete and g preserves countable joins. From the fact that g preserves countable joins, one may deduce that $(\mathcal{N}, \nu)$ is countably additive. It then follows from the first theorem in section 1 that $(\mathcal{N}, \nu)$ is a probability algebra. ▨

Theorem. Suppose $(\mathcal{M}, \mu)$ is a measure algebra. Then there exists a probability algebra $(\mathcal{N}, \nu)$ and a Boolean homomorphism h: $\mathcal{M} \to \mathcal{N}$ such that $\mu = \nu \circ h$.

Proof: The theorem follows directly from the constructions in the two lemmas. For setting $h = g \circ f$, we have $\mu = \delta^* \circ f = \nu \circ g \circ f = \nu \circ h$. ▨

In the proof of the second lemma above, we took a σ-field of subsets that had a countably additive measure and divided it by the ideal consisting of those of its elements with zero measure. As we saw, such a process necessarily results in a probability algebra. With the help of the Loomis-Sikorski Representation Theorem, it is easily shown that any probability algebra can be represented as such a quotient.

Theorem. Suppose $(\mathcal{M}, \mu)$ is a probability algebra. Then there exists a set $\mathcal{S}$, a σ-field $\mathcal{F}$ of subsets of $\mathcal{S}$, a countably additive measure ν on $\mathcal{F}$, and an isomorphism i of $\mathcal{M}$ onto the quotient of $\mathcal{F}$ by the σ-ideal of sets of measure zero such that $\nu(F) = \mu(M)$ whenever $F \in i(M)$.

Proof: The Loomis-Sikorski Representation Theorem supplies us with a σ-field $\mathcal{F}$ of subsets of a set $\mathcal{S}$, σ-ideal I of $\mathcal{F}$ and an isomorphism i of $\mathcal{M}$ onto $\mathcal{F}/I$. Hence we need only verify that the function $\nu: \mathcal{F} \to [0, 1]$ defined by $\nu(F) = \mu(M)$ whenever $F \in i(M)$ is countably additive measure and that I consists precisely of the sets F for which $\nu(F) = 0$.

The second part is easy: the sets F for which $\nu(F) = 0$ are precisely those in $i(\wedge_{\mathcal{M}}) = \wedge \mathcal{F}/I = I$. On the other hand, $\mathcal{S} \in i(\vee_{\mathcal{M}})$, so $\nu(\mathcal{S}) = 1$. Hence we need only show countable additivity for ν. But the canonical homomorphism f: $\mathcal{F} \to \mathcal{F}/I$ is σ-complete. So if we take any disjoint sequence $S_1, S_2, \ldots$ of elements of $\mathcal{F}$, we have $\nu(\cup S_i) = \mu\left(i^{-1}\left(f(\cup S_i)\right)\right) = \mu\left(i^{-1}\left(\vee f(S_i)\right)\right) = \mu\left(\vee i^{-1}\left(f(S_i)\right)\right)$ $= \sum \mu\left(i^{-1}\left(f(S_i)\right)\right) = \sum \nu(S_i)$. ▨

3. Standard Representations for Belief Functions

It follows from the preceding theorem that any belief function can be represented by an allocation into a probability algebra. Suppose, indeed, that we have a belief function Bel: $\mathcal{A} \to [0, 1]$, a measure algebra $(\mathcal{M}, \mu)$ and an allocation $\rho_o: \mathcal{A} \to \mathcal{M}$ such that Bel $= \mu \circ \rho_o$. Then using the probability algebra $(\mathcal{N}, \nu)$ and the Boolean homomorphism h: $\mathcal{M} \to \mathcal{N}$ supplied by our theorem, we may set $\rho = h \circ \rho_o$. The mapping $\rho: \mathcal{A} \to \mathcal{N}$ will then be an allocation into the probability algebra $(\mathcal{N}, \nu)$ and it will represent Bel; for Bel $= \mu \circ \rho_o = \nu \circ h \circ \rho_o = \nu \circ \rho$.

In the sequel, I will generally mean an allocation into a probability algebra whenever I use the term "allocation of probability." When confusion is possible, I will use the word <u>standard</u> to specifically refer to allocations into probability algebras. I will say that an allocation into a probability algebra is a <u>standard allocation</u>, and I will say that it is a <u>standard representation</u> of the belief function it represents.

As we will see, the existence of standard representations will often facilitate our thinking about allocations of probability and belief functions.

It should be noted that when $\rho: \mathcal{A} \to \mathcal{M}$ is a standard representation for the belief function Bel: $\mathcal{A} \to [0,1]$ and $\mathcal{A}_o$ is a subalgebra of $\mathcal{A}$, $\rho|\mathcal{A}_o$ will be a standard representation for the belief function Bel$|\mathcal{A}_o$ on $\mathcal{A}_o$.

4. Quotients of Probability Algebras

One of our fundamental conventions is that the measure of our total probability mass should equal one. It sometimes happens, though, that we want to discard a given probability mass and to regard the probability mass that is left over as our total probability; in this circumstance the measure of our total probability will decrease unless we "renormalize" it. In this section I will briefly describe the process of discarding a probability mass and renormalizing the measure of the remainder.

Essentially, to discard a probability mass means (1) to put the null probability mass in its place and (2) to deduct its contribution from every probability mass to which it contributed. In symbols, the discarding of the probability mass M from a probability algebra $(\mathcal{M}, \mu)$ involves replacing every probability mass $M' \epsilon \mathcal{M}$ by $M' \wedge \overline{M} = M' - M$. Or, to put it a different way, it means identifying all pairs M', M'' of probability masses in $\mathcal{M}$ for which $M' - M = M'' - M$.

But this is precisely what is done when $\mathcal{M}$ is divided by the principal ideal I generated by M. For under that division M' goes into the equivalence class $\left\{M'' \mid M' \triangle M'' \epsilon I\right\} = \left\{M'' \mid M'' \triangle M' \leqslant M\right\} = \left\{M'' \mid M'' - M = M' - M\right\}$. Hence discarding a probability mass means dividing by a principal ideal.

Denote by f the canonical homomorphism of $\mathcal{M}$ onto $\mathcal{M}/I$. Then what measure should be assigned to a given element $f(M') \epsilon \mathcal{M}/I$? Well, $M' = (M' - M) \vee (M' \wedge M)$ and $M' \wedge M$ is being discarded; so $M' - M$ is what is left of M', and it would be natural to adopt $\mu(M' - M)$ as the measure of f(M'). But this procedure will result in a measure of $\mu(\vee_{\mathfrak{M}} - M) = \mu(\overline{M}) = 1 - \mu(M)$ for the unit $\vee_{\mathfrak{M}}/I = f(\vee_{\mathfrak{M}})$. If $\mu(\overline{M}) > 0$ -- i.e., if $M \neq \perp_{\mathfrak{M}}$, then this conflicts with the requirement that the measure of $\vee_{\mathfrak{M}}/I$ should be one. We can correct this difficulty by multiplying all the quantities $\mu(M' - M)$ by a constant in order to increase the measure of $\vee_{\mathfrak{M}}/I$ to one. The appropriate constant is, of course, $1/(1-\mu(M))$. In other words, we define a measure ν on $\mathfrak{M}/I$ by

$$\nu(f(M')) = \frac{1}{1 - \mu(M)} \mu(M' - M).$$

It is easily verified that this is indeed a measure on $\mathfrak{M}/I$. In fact, $(\mathfrak{M}/I, \nu)$ is a probability algebra, provided only that $M \neq \vee_{\mathfrak{M}}$. In the sequel, I will refer to $(\mathfrak{M}/I, \nu)$ as the probability algebra obtained from $(\mathfrak{M}, \mu)$ by discarding M.

5. Orthogonal Sum of Probability Algebras

As I mentioned above, if $(\mathfrak{M}, \mu)$ is a probability algebra and $\mathfrak{N}$ is a complete subalgebra of $\mathfrak{M}$, then $(\mathfrak{N}, \mu|\mathfrak{N})$ is a probability algebra and is said to be embedded in $(\mathfrak{M}, \mu)$. Now suppose that $\mathfrak{M}_1, \ldots, \mathfrak{M}_n$ are independent complete subalgebras of $\mathfrak{M}$. Then they are said to be <u>orthogonal</u> if

$$\mu(M_1 \wedge \ldots \wedge M_n) = \mu(M_1) \ldots \mu(M_n)$$

whenever $M_i \epsilon \mathfrak{M}_i$ for each i, i = 1, ..., n.

In the sequel we will sometimes deal with a collection of probability algebras that are conceived of as having nothing to do with one another and yet which we wish to embed as orthogonal subalgebras of a single overall probability algebra. In this section, we will see how this can be done.

Definition. Suppose $(\mathcal{M}_1, \mu_1)$, ..., $(\mathcal{M}_n, \mu_n)$, $(\mathcal{M}, \mu)$ are probability algebras and that for each i, i = 1, ..., n, $f_i : \mathcal{M}_i \to \mathcal{M}$ is a complete isomorphism into with $\mu_i = \mu \circ f_i$. Suppose further that $f_1(\mathcal{M}_1)$, ..., $f_n(\mathcal{M}_n)$ are independent and orthogonal subalgebras of $\mathcal{M}$. Then $(f_1, \ldots, f_n; (\mathcal{M}, \mu))$ is called an orthogonal sum of $(\mathcal{M}_1, \mu_1)$, ..., $(\mathcal{M}_n, \mu_n)$.

The rest of this section is devoted to showing that an orthogonal sum exists for any finite collection of probability algebras. This will be done by appealing to the construction of "product measures" in measure theory. In particular, I will appeal to the following theorem, which is a long-winded version of the assertion that product measures exist:

Theorem. Let $(\mathcal{S}_1, \mathcal{F}_1, \nu_1)$, ..., $(\mathcal{S}_n, \mathcal{F}_n, \nu_n)$ be "measure spaces." In other words, for each i, i = 1, ..., n, $\mathcal{F}_i$ is a σ-field of subsets of the set $\mathcal{S}_i$ and ν_i is a countably additive measure on $\mathcal{F}_i$. Denote by $\mathcal{S}$ the Cartesian product $\mathcal{S}_1 \times \ldots \times \mathcal{S}_n$. And for each i, i = 1, ..., n, define a mapping $k_i : \mathcal{F}_i \to P(\mathcal{S})$ by setting $k_i(A) = \mathcal{S}_1 \times \ldots \times \mathcal{S}_{i-1} \times A \times \mathcal{S}_{i+1} \times \ldots \times \mathcal{S}_n$. Let $\mathcal{F}$ be the σ-field of subsets of $\mathcal{S}$ generated by $k_1(\mathcal{F}_1) \cup \ldots \cup k_n(\mathcal{F}_n)$. Then

(i) for each i, i = 1, ..., n, k_i is a σ-complete Boolean isomorphism of $\mathcal{F}_i$ into $\mathcal{F}$, and

(ii) there exists a unique countably additive measure ν on $\mathcal{F}$ such that $\nu_i = \nu \circ k_i$ for all i and

$$\nu(A_1 \cap \ldots \cap A_n) = \nu(A_1) \ldots \nu(A_n)$$

whenever $n \geq 1$ and $A_i \in k_i(\mathcal{F}_i)$ for each i, $i = 1, \ldots, n$.

This theorem is proven, for example, in section 37 of Halmos' Measure Theory.

Suppose now that we begin with a collection $(\mathcal{M}_1, \mu_1), \ldots, (\mathcal{M}_n, \mu_n)$ of probability algebras and that we wish to construct an orthogonal sum. Then by the last theorem in section 2, we can suppose that for each i, $i = 1, \ldots, n$, there exists a set $\mathcal{S}_i$, a σ- field $\mathcal{F}_i$ of subsets of $\mathcal{S}_i$, a countably additive measure ν_i on $\mathcal{F}_i$, and an isomorphism j_i of $\mathcal{M}_i$ onto the quotient $\mathcal{F}_i/I_i$, where I_i is the σ - ideal of sets of measure zero and $\nu_i(F) = \mu_i(M)$ whenever $F \in j_i(M)$. Suppose, then, that we let $(\mathcal{S}, \mathcal{F}, \nu)$ and $k_1, \ldots, k_n$ be as in the preceding theorem. Then denoting by I the σ-ideal of $\mathcal{F}$ consisting of all sets of measure zero, we may set $\mathcal{M} = \mathcal{F}/I$ and let μ be the measure on $\mathcal{M}$ inherited from the measure ν on $\mathcal{F}$. Then $(\mathcal{M}, \mu)$ will be a probability algebra and a candidate as an orthogonal sum of $(\mathcal{M}_1, \mu_1), \ldots, (\mathcal{M}_n, \mu_n)$. But we still require the embeddings $f_1, \ldots, f_n$.

First we must use the isomorphisms $k_i: \mathcal{F}_i \to \mathcal{F}$ to construct isomorphisms $k_i': \mathcal{F}_i/I_i \to \mathcal{F}/I$. It is easily seen that whenever A, B $\in \mathcal{F}_i$ differ only by a set of measure zero, their images $k_i(A)$ and $k_i(B)$ differ only by a set of measure zero. Hence k_i' may be defined by setting $k_i'([E]) = [k_i(E)]$. It is easily verified that the k_i' defined in this way are indeed isomorphisms into. Finally, setting $f_i = k_i' \circ j_i$ for each i, $i = 1, \ldots, n$, we obtain the desired embeddings.

6. Bibliographic Notes

With the exception of the ideas in section 3, most of the material in this chapter is fairly well known to students of Boolean algebra. But it is not as widely accessible as the material of the preceding chapter. Several of the proofs in sections 1 and 2 can be gleaned from pp. 55-68 of Halmos' Lectures on Boolean Algebras, but for others I have been unable to find any references.

For a proof of Caratheodory's Extension Theorem, the reader may consult Robert Bartle's Theory of Integration, pp. 98-104.

Another method of proving the main theorem of section 2 would be to take the quotient first and then embed the resulting positive measure algebra in a probability algebra by completing the metric space given by the distance $d(A, B) = \mu(A \Delta B)$. This approach is spelled out in Demetrios A. Kappos' Probability Algebras and Stochastic Spaces, p. 12 and pp. 16-28.

CHAPTER 5. CONDENSABLE ALLOCATIONS

An allocation of probability on a power set $\mathcal{P}(\mathcal{S})$ is condensable if its upper probability function P* obeys

$$P^*(A) = \sup \{P^*(B) \mid B \subset A;\ B \text{ is finite}\}.$$

Condensability is a very important property. It is a property that can generally be expected for belief functions based on empirical evidence; and belief functions that are condensable are intuitively much more transparent than belief functions in general.

This chapter is devoted to the mathematical and intuitive aspects of condensability, and aims at an understanding of the commonality numbers, which provide the best way of describing condensable allocations.

1. Condensability

The theory of degrees of belief set out in the preceding chapters is really built on a single simple intuition: if a given portion of our belief is committed to both of two propositions A and B, then it should be committed to the conjunction $A \wedge B$. It has been my claim that this intuition practically imposes itself -- that a probability mass's being committed to both of two propositions can only mean its being committed to their conjunction.

But one who finds this perception convincing is not likely to stop with pairs of propositions; instead, he is likely to apply the idea to larger, even to infinite collections of propositions. In other words, he will insist that if a given probability mass M is committed to each of a collection $\mathcal{B}$

of propositions, then it must be committed to the logical conjunction of all the elements of $\mathcal{B}$.

If we begin with an arbitrary Boolean algebra of propositions, $\mathcal{A}$, there is no guarantee that $\mathcal{A}$ will contain an element corresponding to the logical conjunction of a given infinite collection of propositions $\mathcal{B} \subset \mathcal{A}$. But suppose that $\mathcal{A}$ can be thought of as the power set of a set $\mathcal{S}$ of possible states of nature, so that a given proposition A in $\mathcal{A}$ asserts that the true state of nature is one of those in a certain subset A of $\mathcal{S}$. Then for any collection $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$, the set-theoretic intersection $\cap \mathcal{B}$ must be interpreted as the logical conjunction of the propositions in $\mathcal{B}$; it says that the true state of nature is in all the sets $B \in \mathcal{B}$, i. e., in their intersection $\cap \mathcal{B}$. In this case, our intuition tells us that a probability mass that is constrained to all the elements of $\mathcal{B}$ should also be constrained to $\cap \mathcal{B}$.

This intuition goes beyond the intuition we have used thus far, and not all allocations of probability on a power set will satisfy it; our rules for allocations imply it for finite collections $\mathcal{B}$, but not for infinite ones. So those allocations that do meet this intuition deserve a special name: A standard allocation of probability $\rho : \mathcal{P}(\mathcal{S}) \rightarrow \mathcal{M}$ over a set $\mathcal{S}$ will be called <u>condensable</u> if for each $M \in \mathcal{M}$ and $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$, M is constrained to $\cup \mathcal{B}$ if and only if it is constrained to each element $B \in \mathcal{B}$.

The requirement that ρ must be standard should not be overlooked; it means that the properties of condensable allocations depend on our intuition about what our probability itself looks like, as well as upon our intuitive understanding of the logical structure of $\mathcal{P}(\mathcal{S})$. In fact, though, condensability is a property of the belief function or the upper probability function and does not depend on which standard representation is used.

Theorem. Suppose $\rho : \mathcal{P}(\mathcal{S}) \to \mathfrak{M}$ is an allocation into the probability algebra $(\mathfrak{M}, \mu)$. Denote by ζ the allowment $\zeta : \mathcal{P}(\mathcal{S}) \to \mathfrak{M} : A \rightsquigarrow \overline{\rho(\tilde{A})}$, by Bel the belief function $\mu \circ \rho$, by P^* the upper probability $\mu \circ \zeta$, and by ct the constraint relation defined by "A ct M if and only if $M \le \rho(A)$. Then the following seven conditions are all equivalent.

(i) ρ is condensable -- i.e., if $\mathcal{B}$ is a non-empty subset of $\mathcal{P}(\mathcal{S})$, $M \in \mathfrak{M}$, and M ct B for all $B \in \mathcal{B}$, then M ct $\cap \mathcal{B}$.

(ii) (ii) $\rho(\cap \mathcal{B}) = \bigwedge_{B \in \mathcal{B}} \rho(B)$ for all non-empty $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$.

(iii) $\zeta(\cup \mathcal{B}) = \bigvee_{B \in \mathcal{B}} \zeta(B)$ for all non-empty $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$.

(iv) For every non-empty $A \subset \mathcal{S}$, there exists a sequence $s_1, s_2, \ldots$ of elements of A and a countable disjoint partition $M_1, M_2, \ldots$ of $\zeta(A)$ such that $M_i \le \zeta(\{s_i\})$ for each positive integer i.

(v) For each $A \subset \mathcal{S}$, $P^*(A) = \sup_{\substack{A' \subset S \\ A' \text{ finite}}} P^*(A')$.

(vi) If $\mathcal{B}$ is an upward net in $\mathcal{P}(\mathcal{S})$, then $P^*(\cup \mathcal{B}) = \sup_{B \in \mathcal{B}} P^*(B)$

(vii) If $\mathcal{B}$ is a downward net in $\mathcal{P}(\mathcal{S})$, then $\mathrm{Bel}(\cap \mathcal{B}) = \inf_{B \in \mathcal{B}} \mathrm{Bel}(B)$.

Proof: (i) $\Rightarrow$ (ii). Since allocations are isotone, $\rho(\cap \mathcal{B}) \le \rho(B)$ for all $B \in \mathcal{B}$, and hence $\rho(\cap \mathcal{B}) \le \bigwedge_{B \in \mathcal{B}} \rho(B)$. On the other hand, $\bigwedge_{B \in \mathcal{B}} \rho(B) \le \rho(B)$ for all $B \in \mathcal{B}$ -- i.e., $\bigwedge_{B \in \mathcal{B}} \rho(B)$ ct B for all $B \in \mathcal{B}$. So by condensability, $\bigwedge_{B \in \mathcal{B}} \rho(B)$ ct $\cap \mathcal{B}$ -- i.e., $\bigwedge_{B \in \mathcal{B}} \rho(B) \le \rho(\cap \mathcal{B})$.

(ii) $\Rightarrow$ (iii). $\zeta(\cup \mathcal{B}) = \overline{\rho(\widetilde{\cup \mathcal{B}})} = \overline{\rho(\widetilde{\bigcup_{B \in \mathcal{B}} B})}$

$= \overline{\rho(\bigcap_{B \in \mathcal{B}} \tilde{B})} = \overline{\bigwedge_{B \in \mathcal{B}} \rho(\tilde{B})} = \bigvee_{B \in \mathcal{B}} \overline{\rho(\tilde{B})} = \bigvee_{B \in \mathcal{B}} \zeta(B).$

(iii) $\Rightarrow$ (iv). For every non-empty $B \subset \mathcal{S}$, $\zeta(B) = \bigvee_{s \in B} \zeta(\{s\})$. Hence (iv) follows by the second theorem of Chapter 4, section 1.

(iv) $\Rightarrow$ (v). We can suppose A is non-empty, and in that case we can choose a sequence $s_1, s_2, \ldots$ of points of A such that $\vee \zeta(\{s_i\}) = \zeta(A)$. But $\bigvee_{i=1}^{\infty} \zeta(\{s_i\}) = \bigvee_{n=1}^{\infty} \zeta(\{s_1, \ldots, s_n\})$, and $\zeta(\{s_1\}) \leq \zeta(\{s_1, s_2\}) \leq \zeta(\{s_1, s_2, s_3\}) \leq \ldots$ is an increasing sequence in $\mathcal{M}$. Hence, by the third theorem of Chapter 4, section 1, $P^*(A) = \mu(\zeta(A)) = \mu(\vee \zeta(\{s_1, \ldots, s_n\})) = \sup_n \mu(\zeta(\{s_1, \ldots, s_n\})) = \sup_n P^*(\{s_1, \ldots, s_n\}) \leq \sup_{\substack{A' \subset A \\ A' \text{ finite}}} P^*(A')$. The inequality $\sup_{\substack{A' \subset A \\ A' \text{finite}}} P^*(A') \leq P^*(A)$ follows, of course, from the monotonicity of P^*.

(v) $\Rightarrow$ (vi). Suppose $\mathcal{B}$ is an upward net and A is a finite subset of $\cup\mathcal{B}$. Then it is easily verified by induction on the number of elements of A that there exists an element $B \in \mathcal{B}$ such that $A \subset B$. Hence if $\mathcal{B}$ is an upward net, $\sup_{B \in \mathcal{B}} P^*(B) \geq \sup_{\substack{A \subset \cup\mathcal{B} \\ A \text{ finite}}} P^*(A) = P^*(\cup \mathcal{B})$. The inequality $\sup_{B \in \mathcal{B}} P^*(B) \leq P^*(B)$ follows, of course, from the monotonicity of P^*.

(vi) $\Rightarrow$ (vii). Suppose $\mathcal{B}$ is a downward net in $P(\mathcal{S})$. Then $\mathcal{C} = \{\tilde{B} | B \in \mathcal{B}\}$ is an upward net in $P(\mathcal{S})$. Hence

$$\mathrm{Bel}(\cap\mathcal{B}) = 1 - P^*(\widetilde{\cap\mathcal{B}}) = 1 - P^*(\cup\mathcal{C}) = 1 - \sup_{C \in \mathcal{C}} P^*(C)$$

$$= 1 - \sup_{B \in \mathcal{B}} P^*(\tilde{B}) = \inf_{B \in \mathcal{B}} (1 - P^*(\tilde{B})) = \inf_{B \in \mathcal{B}} \mathrm{Bel}(B).$$

(vii) $\Rightarrow$ (ii). Suppose $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$ is non-empty. Then $\mathcal{C} = \{\cap\mathcal{A} | \mathcal{A} \subset \mathcal{B},\ \mathcal{A} \text{ finite}\}$ is a downward net in $\mathcal{P}(\mathcal{S})$. But $\cap\mathcal{B} = \cap\mathcal{C}$ and $\bigwedge_{C \in \mathcal{C}} \rho(C) = \bigwedge_{B \in \mathcal{B}} \rho(B)$. Hence

$$\mu(\rho(\cap\mathcal{B})) = \mu(\rho\cap\mathcal{C})) = \mathrm{Bel}(\cap\mathcal{C}) = \inf_{C \in \mathcal{C}} \mathrm{Bel}(C)$$

$$= \inf_{C \in \mathcal{C}} \mu(\rho(C)) = \mu(\bigwedge_{C \in \mathcal{C}} \rho(C)) = \mu(\bigwedge_{B \in \mathcal{B}} \rho(B)).$$

Since $\rho(\cap \mathcal{B}) \subset \bigwedge_{B \in \mathcal{B}} \rho(B)$ and μ is positive, it follows that $\rho(\cap \mathcal{B}) = \bigwedge_{B \in \mathcal{B}} \rho(B)$.

(ii) $\Rightarrow$ (i) If M ct B for all $B \in \mathcal{B}$, then $M \subset \rho(B)$ for all $B \in \mathcal{B}$, or $M \subset \bigwedge_{B \in \mathcal{B}} \rho(B)$. Hence $M \subset \rho(\cap \mathcal{B})$, or M ct $\cap \mathcal{B}$. ▨

Since conditions (v), (vi) and (vii) make no reference to any particular standard representation for the belief function or upper probability function, this theorem justifies the assertion that condensability is a property of the belief function or upper probability function and does not depend on which standard representation is used. More generally, the theorem shows that the adjective <u>condensable</u> can properly be applied to the constraint relation, the allowment, the upper probability function or the belief function, as well as to the allocation ρ. I will follow such a usage in the sequel.

Condition (iv) is of particular interest for the intuitive understanding of condensability. It states that the probability mass $\zeta(A)$ -- the total probability mass that can get into B -- can be divided into a countable number of discrete pieces, each of which can get into some single point of B. We will shortly see why this property deserves to be called "condensability."

It is condition (v) that we will deal with most often in the sequel. Its utility is obvious -- it means that the entire upper probability function is determined by its values on finite subsets and thus allows us to examine the structure of condensable upper probability functions much more closely. We will begin this closer examination in section 3.

In my definition of condensability, I have required that the allocation or belief function be on a power set $\mathcal{P}(\mathcal{S})$. This may seem unnecessarily restrictive, for the definition could easily be extended to any complete Boolean algebra in which arbitrary meets and joins can be understood as conjunctions and disjunctions. It is not clear, however, that there are any such Boolean algebras which are not isomorphic to power sets; and hence it is not clear whether the seemingly more general formulation is of any real interest. In any case, the upper probability functions that we will be concerned with will be on power sets.

There are many ways in which condensable belief functions are more attractive than belief functions in general. Consider, for example, the problem of sets of "upper probability zero." If the upper probability function $P^*: \mathcal{P}(\mathcal{S}) \longrightarrow [0,1]$ is condensable, then the set

$$S = \bigcup \{ S' \mid P^*(S') = 0 \}$$

will obey $P^*(S) = 0$. (This follows from condition (vi) in the preceding theorem.) The significance of this fact is that it makes it possible to interpret "$P^*(S) = 0$" as really meaning that the upper probability function P^* holds S to be impossible. In the case of non-condensable belief functions -- for example, in the case of "continuous" probability functions -- such an interpretation is, somewhat paradoxically, impossible.

2. Mobile Probability Masses

A condensable allocation on a power set $\mathcal{P}(\mathcal{S})$ can be interpreted in a very vivid way if we think of the set $\mathcal{S}$ geometrically and think of our probability as being distributed over it. More precisely, let us think of our probability not as being distributed in a fixed way, but rather as having a certain degree of mobility. In other words, the various probability masses in $\mathcal{M}$ are to be allowed to move around, to some extent, within $\mathcal{S}$.

The extent of the mobility is specified by the constraint relation ct between $\mathcal{M}$ and $\mathcal{P}(\mathcal{S})$; if a probability mass $M \in \mathcal{M}$ is constrained to a set $A \subset \mathcal{S}$, this means precisely that neither M nor any subelement of M can get out of A. A glance at the rules for constraint relations in section 2 of Chapter 4 will reveal that those rules are all immediately obvious from this geometric picture. And the condition of condensability is equally obvious; for if all of a probability mass is constrained to stay inside A for each A in some subset $\mathcal{B}$ of $\mathcal{P}(\mathcal{S})$, then it must be constrained to stay inside $\bigcap\mathcal{B}$.

An even more vivid understanding of condensability can be obtained from condition (iv) of the theorem in the preceding section. Intuitively, this condition means that though the constraints on the probability mass $\zeta(A)$ might allow it to become spread out over A in a completely diffuse fashion (as in the case of a "continuous distribution" of probability), it must always be possible to condense it into a collection of discrete pieces, just as a diffuse mass of water vapor can be condensed into a collection of drops. The word "condensability" is meant to bring to mind the possibility of such a condensation.

It is easy to think about a subset A's degree of belief Bel(A) and upper probability P*(A) in terms of this picture. Bel(A) is simply the amount of probability that cannot get out of A, while P*(A) is the amount of probability that can get into A.

If we concentrate on a probability mass $M \in \mathcal{M}$, it is natural to ask just how constrained M is. Evidently there will be a whole, possibly quite large, set $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$ of subsets of $\mathcal{S}$ to which M is constrained. By condensability, M will also be constrained to $\cap\mathcal{B}$, and this will be the smallest region to which all of it is constrained -- its "tightest" constraint. But as we saw in section 9 of Chapter 2, the existence of such a "tightest" constraint for each probability mass can be described by saying that there exists a "constraint mapping" $\lambda: \mathcal{M} \to \mathcal{P}(\mathcal{S})$ that maps each probability mass to its tightest constraint. So condensability has to do with the existence of a constraint mapping.

This may be puzzling, for in Chapter 2 we saw that any belief function can be represented by an allocation of probability for which a constraint mapping exists. But the allocation constructed there was not necessarily standard -- it was into a "measure algebra" but not necessarily into a "probability algebra." And when the allocation is extended to one into a probability algebra, the constraint mapping may be lost. In fact, it will be unless the belief function is condensable.

<u>Theorem</u>. Suppose $\rho: \mathcal{P}(\mathcal{S}) \to \mathcal{M}$ is a standard allocation of probability. Then ρ is condensable if and only if a constraint mapping $\lambda: \mathcal{M} \to \mathcal{P}(\mathcal{S})$ exists for ρ.

<u>Proof</u>: If ρ is condensable, then the mapping $\lambda: \mathcal{M} \to \mathcal{P}(\mathcal{S}): M \rightsquigarrow \cap\{A \mid A \subset \mathcal{S},\ M \text{ ct } A\}$ is a constraint mapping for ρ. If a constraint mapping $\lambda: \mathcal{M} \to \mathcal{P}(\mathcal{S})$ exists, then M ct A if and only $\lambda(M) \subset A$; so if M ct B for all $B \in \mathcal{B}$ it follows that $\lambda(M) \subset \mathcal{B}$ for all $B \in \mathcal{B}$ and $\lambda(M) \subset \cap\mathcal{B}$, whence M ct $\cap\mathcal{B}$. ▨

3. Upper Probabilities for Finite Subsets

A condensable upper probability function $P^*: \mathcal{P}(\mathcal{S}) \to [0, 1]$ is determined by its values on finite subsets of $\mathcal{S}$. Denoting by $\mathcal{F}(\mathcal{S})$ the set of all finite subsets of $\mathcal{S}$, we can express this by saying that P^* is completely determined by $P_o^*: \mathcal{F}(\mathcal{S}) \to [0,1]$, where $P_o^* = P^* \mid \mathcal{F}(\mathcal{S})$.

This fact leads us naturally to inquire about the properties of P_o^*. On the one hand, we might ask what properties P_o^* will have on account of P^*'s being a condensable upper probability function; and on the other hand, we might look for conditions on a function $f: \mathcal{F}(\mathcal{S}) \to [0, 1]$ that are sufficient to assure that the function $P^*: \mathcal{P}(\mathcal{S}) \to [0, 1]: A \rightsquigarrow \sup_{\substack{A' \subset A \\ A' \text{finite}}} f(A)$ should be a condensable upper probability function. The following lemma will help us state such conditions:

<u>Lemma.</u> Let f be a real function on the set $\mathcal{F}(\mathcal{S})$ of all finite subsets of a non-empty set $\mathcal{S}$, and denote

$$\nabla_n^f (B; A_1, \ldots, A_n) = \sum_{J \subset \{1, \ldots, n\}} (-1)^{\text{card } J} f(B \cup (\bigcup_{i \in J} A_i))$$

whenever $n \geq 1$ and $B, A_1, \ldots, A_n \in \mathcal{F}(\mathcal{S})$. Now fix $A_1, \ldots, A_n$, and for each i, $i = 1, \ldots, n$, set

$$A_i = \{a_{i1}, \ldots, a_{ik_i}\}$$

and for each j, $j = 0, 1, \ldots, k_i$, set

$$A_i^j = \{a_{i1}, \ldots, a_{ij}\}.$$

($A_i^o = \emptyset$ for all i, $i = 1, \ldots, n$.) Then

$$\nabla_n(B; A_1, \ldots, A_n) = \sum\left\{\nabla_n\left(B \cup A_1^{j_1-1} \cup \cdots \cup A_n^{j_n-1};\right.\right.$$

$$\left.\left.\{a_{1j_1}\}, \ldots, \{a_{nj_n}\}\right) \middle| 1 \le j_i \le k_i\right\}.$$

Proof: If $k_i = 0$ for some i, then $A_i = \phi$, and it is evident from the fact that ∇_n is a successive difference (cf. Chapter 1, section 3) that $\nabla_n(B, A_1, \ldots, A_n) = 0$; on the other hand, the right-hand side above would also be zero, for there would be no terms in the summation. Hence we may assume that $k_i > 0$ for $i = 1, \ldots, n$.

In that case,

$$\text{r.h.s.} = \sum_{\substack{(j_1, \ldots, j_n), \\ \phi \le j_i \le k_i}} \sum_{J \subset \{1, \ldots, n\}} (-1)^{\text{card } J} f\left(B \cup (\bigcup_{i \in J} A_i^{j_i}) \cup (\bigcup_{i \notin J} A_i^{j_i-1})\right)$$

$$= \sum_{\substack{(j_1, \ldots, j_n), \\ 0 \le j_i \le k_i}} f(B \cup (\bigcup_{i=1}^{n} A_i^{j_i})) \sum\left\{(-1)^{\text{card } J} \middle| \{i \mid j_i = k_i\} \subset J \subset \{i \mid j_i \neq 0\}\right\}$$

$$= \sum_{\substack{(j_1, \ldots, j_n), \\ j_i = 0 \text{ or } k_i \text{ for each } i}} f(B \cup (\bigcup_{i=1}^{n} A_i^{j_i}))\ (-1)^{\#\text{ of } i \text{ for which } j_i = k_i}$$

$$= \sum_{J \subset \{1, \ldots, n\}} f(B \cup (\bigcup_{i \in J} A_i))\ (-1)^{\text{card } J}$$

$$= \nabla_n(B; A_1, \ldots, A_n).$$

▨

Theorem. Suppose f is a real function on $\mathcal{F}(\mathcal{S})$, the set of all finite subsets of a non-empty set $\mathcal{S}$. Then the real function P^* on $\mathcal{P}(\mathcal{S})$ defined by $P^*(A) = \sup_{\substack{A' \subseteq A \\ A' \text{ finite}}} f(A')$ is a condensable upper probability function if and only if

(i) $f(\emptyset) = 0$

(ii) $\sup_{A \in \mathcal{F}(\mathcal{S})} f(A) = 1$

(iii) If $A, B \in \mathcal{F}(\mathcal{S})$ and $A \neq \emptyset$, then $\sum_{T \subset A} (-1)^{\text{card } T} f(B \cup T) \leq 0.$

Proof: First we must show that if $P^*: \mathcal{P}(\mathcal{S}) \to [0,1]$ is a condensable upper probability function, then $f = P^* | \mathcal{F}(\mathcal{S})$ satisfies the three conditions. But (i) and (ii) are obvious. Now we may write $A = \{s_1, \ldots, s_n\}$ for some $n \geq 1$, and $\sum_{T \subset B} (-1)^{\text{card } T} f(B \cup T)$ then becomes $\sum_{J \subset \{1, \ldots, n\}} (-1)^{\text{card } J} P^*(B \cup (\bigcup_{i \in J} \{s_i\})) = \nabla_n^{P^*}(B; \{s_1\}, \ldots, \{s_n\})$, and this is non-positive according to section 3 of Chapter 1.

Next, we must show that P^* is a condensable upper probability function if f satisfies the three conditions and P^* is defined by $P^*(A) = \sup_{\substack{A' \subseteq A \\ A' \text{finite}}} f(A')$. But the relations $P^*(\emptyset) = 0$ and $P^*(\mathcal{S}) = 1$ are evident from (i) and (ii). Hence, by the last theorem of section 3 of Chapter 1, we need only show that $\nabla_n(B; A_1, \ldots, A_n) \leq 0$ for all $B, A_1, \ldots, A_n \in \mathcal{P}(\mathcal{S})$. But we have just seen that $\sum_{T \subset A} (-1)^T (B \cup T) = \nabla_n^f(B; \{s_1\}, \ldots, \{s_n\})$, where $A = \{s_1, \ldots, s_n\}$, so (iii) above asserts that $\nabla_n(B; A_1, \ldots, A_n) \leq 0$ in the case where B is finite and the A_i are singletons. The case where B and the A_i are all finite follows by the lemma.

By the definition of P^*, the values $P^*(A)$ can always be approximated by values $P^*(A')$, where $A' \subseteq A$ and A' is finite, so we can easily establish that $\nabla_n(B; A_1, \ldots, A_n) \leq 0$ in general by approximating each upper probability with the upper probability of a finite subset. Suppose, indeed, that

$$0 < \varepsilon = \nabla_n(B; A_1, \ldots, A_n) = P^*(B) - \Sigma P^*(B \cup A_i) + \cdots \\ + (-1)^n P^*(B \cup A_i, \cup \ldots \cup A_n).$$

Then since there are 2^n terms on the right-hand side of this inequality, we can approximate B, $A_1, \ldots, A_n$ by finite subsets B', $A_1', \ldots, A_n'$ such that $P^*(B \cup A_{i_1} \cup \ldots \cup A_{i_k})$ differs from $P^*(B' \cup A'_{i_1} \cup \ldots \cup A'_{i_k})$ by less than $1/2 \cdot \varepsilon / 2^n$ for each $(i_1, \ldots, i_k)$ such that $1 \leq i_1 \leq \ldots \leq i_k \leq n$. Hence, the quantity $\nabla_n(B', A_1', \ldots, A_n')$ would also be positive, contradicting our conclusion in the preceding paragraph. ▨

<u>Theorem</u>. Suppose $\mathcal{S}$ is a non-empty set, $(\mathcal{M}, \mu)$ is a measure algebra and

$$\zeta_o : \mathcal{S} \to \mathcal{M}$$

is such that for any $\varepsilon > 0$ there exists a finite subset $\{s_1, \ldots, s_n\}$ of $\mathcal{S}$ such that

$$\mu(\zeta_o(s_1) \vee \ldots \vee \zeta_o(s_n)) > 1 - \varepsilon .$$

Then the function $P^*: \mathcal{P}(\mathcal{S}) \to [0, 1]$ defined by $P^*(\phi) = 0$ and

$$P^*(A) = \sup \{\mu(\zeta_o(s_1) \vee \ldots \vee \zeta_o(s_n)) \mid n \geq 1; \{s_1, \ldots, s_n\} \subset A\}$$

for $A \neq \phi$ is a condensable upper probability function.

<u>Proof</u>: Evidently, $P^*(A) = \sup_{\substack{A' \subset A \\ A' \text{finite}}} f(A)$, where $f(\phi) = 0$ and $f(\{s_1, \ldots, s_n\}) = \mu(\zeta_o(s_1) \vee \ldots \vee \zeta_o(s_n))$. And $\sup_{A \in \mathcal{F}(\mathcal{S})} f(A) = 1$. So by the preceding theorem we need only show that if A, $B \in \mathcal{F}(\mathcal{S})$

and $A \neq 0$, then

$$\sum_{T \subset A} (-1)^{\text{card } T} f(B \cup T) \leq 0.$$

Now set $A = \{s_1, \ldots, s_n\}$ and set

$$M = \begin{cases} \Lambda_{\mathcal{M}} & \text{if } B = \emptyset \\ \zeta(t_1) \vee \ldots \vee \zeta(t_m) & \text{if } B = \{t_1, \ldots, t_m\}, \text{ where } m \geq 1. \end{cases}$$

Then

$$\sum_{T \subset A} (-1)^{\text{card } T} f(B \cup T)$$

$$= \mu(M) - \Sigma \mu(M \vee \zeta_o(s_i)) + \Sigma \mu(M \vee \zeta_o(s_i) \vee \zeta_o(s_j))$$

$$- + \ldots + (-1)^{n+1} \mu(M \vee \zeta_o(s_1) \vee \ldots \vee \zeta_o(s_n))$$

$$= \mu(M) - \mu((M \vee \zeta_o(s_1)) \wedge \ldots \wedge (M \vee \zeta_o(s_n)))$$

$$= \mu(M) - \mu(M \vee (\wedge \zeta_o(s_i)))$$

$$\leq 0.$$

////

4. Commonality Numbers

Let $\rho: \mathcal{P}(\mathcal{S}) \longrightarrow \mathcal{M}$ be a condensable allocation of probability, and let ζ be the allowment associated with ρ. In other words, $\zeta(A) = \overline{\rho(\tilde{A})}$. Then for each $s \in \mathcal{S}$, $\zeta(\{s\})$ is the total probability mass that can reach the point s. And for any non-empty finite subset $A = \{s_1, \ldots, s_n\}$ of $\mathcal{S}$, $\zeta(\{s_1\}) \wedge \ldots \wedge \zeta(\{s_n\})$ is the total probability mass that can reach each and every point of A -- i.e., the total probability mass that can move completely freely within A.

Now if A = $\emptyset$, the total probability mass that can reach each and every point of A is $\mathcal{V}_{\mathcal{M}}$. Hence it is natural to define a mapping $\gamma: \mathcal{F}(\mathcal{S}) \rightarrow \mathcal{M}$ by $\gamma(\emptyset) = \mathcal{V}_{\mathcal{M}}$ and $\gamma(\{s_1, \ldots, s_n\} = \zeta(\{s_1\}) \wedge \ldots \wedge \zeta(\{s_n\})$. As it turns out, the measures of the probability masses $\gamma(A)$, $A \in \mathcal{F}(\mathcal{S})$, are very important and hence deserve a name. Setting $Q = \mu \circ \gamma$, where μ is the measure on $\mathcal{M}$, I will call Q(A) the "commonality number" for A, and I will call $Q: \mathcal{F}(\mathcal{S}) \rightarrow [0, 1]$ the "commonality function" associated with ρ.

Notice that the commonality number Q(A) decreases as A is enlarged. Indeed, $Q(\emptyset) = \mu(\mathcal{V}_{\mathcal{M}}) = 1$, and $Q(\{s_1, \ldots, s_n, s_{n+1}\}) = \mu(\zeta(\{s_1\} \wedge \ldots \wedge \zeta(\{s_n\}) \wedge \quad \zeta(\{s_{n+1}\})) \leq \mu(\zeta(\{s_1\}) \wedge \ldots \wedge \zeta(\{s_n\})) = Q(\{s_1, \ldots, s_n\})$. This contrasts sharply with the behavior of the upper probability P*(A) which begins at zero when A = $\emptyset$ and increases as A is enlarged.

Actually, commonality numbers and upper probabilities are related by a much more extensive duality. For while the commonality numbers give the measures of the intersections of the probability masses $\zeta(\{s_i\})$, upper probabilities give the measures of their unions:

$$Q(\{s_1, \ldots, s_n\}) = \mu(\zeta(\{s_1\}) \wedge \ldots \wedge \zeta(\{s_n\})),$$

while

$$P^*(\{s_1, \ldots, s_n\}) = \mu(\zeta(\{s_1, \ldots, s_n\})) = \mu(\zeta(\{s_1\}) \vee \ldots \vee \zeta(\{s_n\})).$$

Now we know from the theory of measure (and from Chapter 1, section 5) that the measures of finite meets can always be expressed in terms of the measures of finite joins and vice-versa:

$$\mu(M_1 \wedge \ldots \wedge M_n) = \Sigma\mu\,(M_i) - \Sigma\mu\,(M_i \vee M_j) + - \ldots + (-1)^{n+1}\mu(M_1 \vee \ldots \vee M_n)$$

and

$$\mu(M_1 \vee \ldots \vee M_n) = \Sigma\mu\,(M_i) - \Sigma\mu\,(M_i \wedge M_j) + - \ldots + (-1)^{n+1}\,\mu(M_1 \wedge \ldots \wedge M_n)$$

for all $M_1, \ldots, M_n \in \mathcal{M}$. So for all non-empty finite subsets $\{s_1, \ldots, s_n\}$ of $\mathcal{S}$,

$$Q(\{s_1, \ldots, s_n\}) = \Sigma\, P^*(\{s_i\}) - \Sigma\, P^*(\{s_i, s_j\} + - \ldots + (-1)^{n+1} P^*(\{s_1, \ldots, s_n\})$$

and

$$P^*(\{s_1, \ldots, s_n\}) = \Sigma\, Q(\{s_i\}) - \Sigma\, Q(\{s_i, s_j\}) + - \ldots + (-1)^{n+1}\, Q(\{s_1, \ldots, s_n\}).$$

It is evident from this last formula that the commonality numbers determine the upper probabilities for finite subsets and hence the entire condensable upper probability function.

So in the condensable case commonality functions are simply another form in which belief functions may be specified. It will be useful to know what properties fully characterize them.

Definition. A real function Q on the set $\mathcal{F}(\mathcal{S})$ of all finite subsets of a non-empty set $\mathcal{S}$ is called a commonality function if

(i) $Q(\emptyset) = 1$,

(ii) $\inf_{A \in \mathcal{F}(\mathcal{S})} \sum_{T \subset A} (-1)^{\text{card } T}\, Q(T) = 0$,

(iii) If $A, B \in \mathcal{F}(\mathcal{S})$, then $\sum_{T \subset B} (-1)^{\text{card } T}\, Q(A \cup T) \geq 0$.

Theorem. If the function Q on $\mathcal{F}(\mathcal{S}))$ is a commonality function, then it takes values in the interval $[0,1]$.

Proof: Setting $B = \{s\}$ in (iii) yields $Q(A) - Q(A \cup \{s\}) \geq 0$, or $Q(A) \geq Q(A \cup \{s\})$ for all $A \in \mathcal{F}(\mathcal{S})$ and $s \in \mathcal{S}$. But $Q(\phi) = 1$. Hence $Q(A) \leq 1$ for all $A \in \mathcal{F}(\mathcal{S})$.

Setting $B = \phi$ in (iii) yields $Q(A) \geq 0$ for all $A \in \mathcal{F}(\mathcal{S})$. ▨

Lemma: Suppose f is a real function on the set $\mathcal{F}(\mathcal{S})$ of finite subsets of a set $\mathcal{S}$. And suppose $A, B \in \mathcal{F}(\mathcal{S})$. Then

$$\sum_{T \subset A \cup B} (-1)^{\text{card } T} f(T) = \sum_{R \subset A} \sum_{S \subset B} (-1)^{\text{card } R} (-1)^{\text{card } S} f(R \cup S).$$

Proof:
$$\sum_{R \subset A} \sum_{S \subset B} (-1)^{\text{card } R} (-1)^{\text{card } S} f(R \cup S)$$

$$= \sum_{T \subset A \cup B} f(T) \sum \{(-1)^{\text{card } R + \text{card } S} \mid R \subset A;\ S \subset B;\ R \cup S = T\}$$

$$= \sum_{T \subset A \cup B} f(T)\ (-1)^{\text{card}(A-B) + \text{card}(B-A)} \sum \{(-1)^{\text{card } R + \text{card } S} \mid R, S \subset A \cap B;\ R \cup S = A \cap B\}.$$

But for any subset A,

$$\sum \{(-1)^{\text{card } R + \text{card } S} \mid R, S \subset A;\ R \cup S = A\} = (-1)^{\text{card } A}.$$

The lemma follows. ▨

Theorem. Suppose $P^*: \mathcal{P}(\mathcal{S}) \rightarrow [0,1]$ is a condensable upper probability function and define the function Q on $\mathcal{F}(\mathcal{S})$ by $Q(\phi) = 1$ and

$$Q(A) = - \sum_{T \subset A} (-1)^{\text{card } T} P^*(T)$$

for non-empty $A \in \mathcal{J}(\mathcal{S})$. Then Q is a commonality function.

<u>Proof</u>: (i) $Q(\phi) = 1$ by definition.

(ii) If $A = \phi$, then $\sum_{T \subset A} (-1)^{\text{card } T} Q(T) = Q(\phi) = 1$.

If, on the other hand, $A \neq \phi$, then we can write $A = \{s_1, \ldots, s_n\}$ with $n \geq 1$, and

$$\sum_{T \subset A} (-1)^{\text{card } T} Q(T) = 1 - \sum_{\substack{T \subset A \\ T \neq \phi}} (-1)^{\text{card } T} \sum_{R \subset T} (-1)^{\text{card } R} P^*(R)$$

$$= 1 - \sum_{R \subset A} (-1)^{\text{card } R} P^*(R) \Big(\sum_{R \subset T \subset A} (-1)^{\text{card } T} \Big)$$

$$= 1 - P^*(A).$$

Hence $\inf_{A \in \mathcal{J}(\mathcal{S})} \sum_{T \subset A} (-1)^{\text{card } T} Q(T) = \inf_{A \in \mathcal{J}(\mathcal{S})} (1 - P^*(A)) = 1 - \sup_{A \in \mathcal{J}(\mathcal{S})} P^*(A) = 0.$

(iii) Finally, we need to show that

$$\sum_{T \subset B} (-1)^{\text{card } T} Q(A \cup T) \geq 0$$

for all $A, B \in \mathcal{J}(\mathcal{S})$. If $A = \phi$, this reduces to

$$\sum_{T \subset B} (-1)^{\text{card } T} Q(T) \geq 0,$$

and we just proved this. Hence we may assume that $A \neq \phi$, writing $A = \{s_1, \ldots, s_n\}$ and $B = \{t_1, \ldots, t_p\}$, where $n \geq 1$ and $p \geq 0$. Then

$$\sum_{T \subset B} (-1)^{\text{card } T} Q(A \cup T) = - \sum_{T \subset B} (-1)^{\text{card } T} \sum_{R \subset A \cup T} (-1)^{\text{card } R} P^*(R)$$

$$= - \sum_{T \subset B} (-1)^{\text{card } T} \sum_{R \subset A} \sum_{S \subset T} (-1)^{\text{card } R} (-1)^{\text{card } S} P^*(R \cup S)$$

$$= - \sum_{R \subset A} (-1)^{\text{card } R} \sum_{S \subset B} (-1)^{\text{card } S} P^*(R \cup S) \left(\sum_{S \subset T \subset B} (-1)^{\text{card } T} \right)$$

$$= - \sum_{R \subset A} (-1)^{\text{card } R} P^*(B \cup R).$$

But $\sum_{R \subset A} (-1)^{\text{card } R} P^*(B \cup R) \leq 0$ by the last theorem of the preceding section. ▨

<u>Theorem</u>. Suppose Q: $\mathcal{F}(\mathcal{S}) \to [0, 1]$ is a commonality function, and define the function P* on $\mathcal{P}(\mathcal{S})$ by $P^*(\phi) = 0$,

$$P^*(A) = - \sum_{\substack{T \subset A \\ T \neq \phi}} (-1)^{\text{card } T} Q(T)$$

for finite non-empty subsets A of $\mathcal{S}$, and

$$P^*(A) = \sup_{\substack{A' \subset A \\ A' \text{finite}}} P^*(A')$$

for infinite subsets $A \subset \mathcal{S}$. Then P* is a condensable upper probability function.

<u>Proof</u>: By the last theorem of the preceding section, it suffices to prove that

(i) $P^*(\phi) = 0$

(ii) $\sup_{A \in \mathcal{F}(\mathcal{S})} P^*(A) = 1$

and (iii) If A, B $\in \mathcal{F}(\mathcal{S})$ and $A \neq \phi$, then $\sum_{T \subset A} (-1)^{\text{card } T} P^*(B \cup T) \leq 0.$

But (i) is given by convention. As for (ii), for finite non-empty subsets A,

$$P^*(A) = 1 - \sum_{T \subset A} (-1)^{\text{card } T} Q(T),$$

so

$$\sup_{A \in \mathcal{F}(\mathcal{S})} P^*(A) = 1 - \inf_{A \in \mathcal{F}(\mathcal{S})} \Sigma (-1)^{\text{card } T} Q(T) = 1.$$

To prove (iii), note that

$$\sum_{T \subset A} (-1)^{\text{card } T} P^*(B \cup T) = - \sum_{T \subset A} (-1)^{\text{card } T} \sum_{\substack{R \subset B \cup T \\ R \neq \emptyset}} (-1)^{\text{card } R} Q(R)$$

$$= - \sum_{T \subset A} (-1)^{\text{card } J} \sum_{\substack{R \subset B \\ S \subset T \\ \text{either } R \text{ or } S \neq \emptyset}} (-1)^{\text{card } R} (-1)^{\text{card } S} Q(R \cup S)$$

$$= - \sum_{R \subset B} (-1)^{\text{card } R} \sum_{\substack{S \subset A \\ S \neq \emptyset \text{ if } R = \emptyset}} (-1)^{\text{card } S} Q(R \cup S) \sum_{S \subset T \subset A} (-1)^{\text{card } T}$$

$$= - \sum_{R \subset B} (-1)^{\text{card } R} Q(A \cup S).$$

But $\sum_{R \subset B} (-1)^{\text{card } R} Q(A \cup S) \geq 0$ by the definition of commonality functions. ▨

In the sequel, we will sometimes examine a real function on $\mathcal{F}(\mathcal{S})$ $-\{\emptyset\}$ with the question as to whether it can be "renormalized" so as to yield a commonality function. In other words, given a function Q_1 on $\mathcal{F}(\mathcal{S}) - \{\emptyset\}$, we will want to know whether there exists a constant K such that the function Q on $\mathcal{F}(\mathcal{S})$ defined by

$$Q(A) = \begin{cases} 1 & \text{if } A = \phi \\ K\,Q_1(A) & \text{if } A \neq \phi \end{cases}$$

is a commonality function. The following theorem gives the conditions under which such a constant does exist.

Theorem. Suppose $\mathcal{S}$ is a non-empty set and Q_1 is a real function on $\mathcal{F}(\mathcal{S}) - \{\phi\}$. And for each positive number K define a real function Q_K on $\mathcal{F}(\mathcal{S})$ by

$$Q_K(A) = \begin{cases} 1 & \text{if } A = \phi \\ K\,Q_1(A) & \text{if } A \neq \phi. \end{cases}$$

Then Q_K is a commonality function if and only if

(i) $$\sup_{A \in (\mathcal{F}(\mathcal{S}) - \{\phi\})} \sum_{\substack{T \subset A \\ T \neq \phi}} (-1)^{1 + \text{card } T}\, Q_1(T) = 1/K,$$

(ii) If A, B $\in (\mathcal{F}(\mathcal{S}) - \{\phi\})$, then

$$\sum_{\substack{T \subset B \\ T \neq \phi}} (-1)^{1+\text{card } T}\, Q_1(A \cup T) \leq 1/K.$$

This theorem follows directly from the definition of commonality functions.

The preceding discussion has been primarily concerned with the relation between Q and P*. The formulae connecting Q and Bel are in some respects simpler and worth recording:

$$Q(A) = \sum_{T \subset A} (-1)^{\text{card } T} \, Bel(\tilde{T})$$

for all $A \in \mathcal{F}(\mathcal{S})$, including ϕ; and

$$Bel(A) = \sum_{A \subset T} (-1)^{\text{card } \tilde{T}} \, Q(\tilde{T})$$

for cofinite A and

$$Bel(A) = \inf_{\substack{A \subset A' \\ A' \text{cofinite}}} \sum_{A' \subset T} (-1)^{\text{card } \tilde{T}} \, Q(\tilde{T})$$

in general. A subset A of $\mathcal{S}$ is said to be <u>cofinite</u> if $\tilde{A} = \mathcal{S} \sim A$ is finite. The quantity

$$\inf_{\substack{A \subset A' \\ A' \text{cofinite}}} \sum_{A' \subset T} (-1)^{\text{card } \tilde{T}} \, Q(\tilde{T})$$

can be thought of intuitively as the summation of $(-1)^{\text{card } T} Q(T)$ over all finite T that do not intersect A.

5. Restricting Condensable Allocations

It is not difficult to prove that a complete subalgebra of a power set is itself isomorphic to a power set. Hence, it makes sense to ask whether a condensable allocation $\rho : \mathcal{P}(\mathcal{S}) \rightarrow \mathcal{M}$ remains condensable when it is restricted to a complete subalgebra $\mathcal{A} \subset \mathcal{P}(\mathcal{S})$.

The answer is obviously yes; for, since

$$\rho(\cap B) = \bigwedge_{B \in \mathcal{B}} \rho(B)$$

holds for all $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$ it will certainly hold for all $\mathcal{B} \subset \mathcal{Q}$.

CHAPTER 6. EXTENSION AND COMBINATION

In this chapter we begin to see just how flexible belief functions are. In particular, we find that belief functions on given Boolean algebras can sometimes be used to obtain belief functions on more complicated Boolean al gebras.

The central concern of the chapter is a rule that enables one to combine belief functions on different Boolean algebras into a single resultant belief function on their independent sum. A quite general rule is adduced for such combination, and a much simpler rule is derived for the condensable case.

The existence of such a rule also leads to the exploration of the notion of subalgebras being "independent" with respect to a belief function. As it turns out, it is convenient to distinguish between the notions of "orthogonality" and "cognitive independence," notions which collapse into a single notion in the case of probability functions.

1. Extending Allocations of Probability

In this section we will study one of the most remarkable and fruitful features of the theory of allocations: the fact that an allocation of probability on a subalgebra of a larger algebra always has a natural extension to the larger algebra. The existence of such an extension results from the fundamental intuition that any portion of our belief that is committed to a given proposition must also be committed to any proposition that it implies -- i.e., to any more inclusive proposition.

Suppose, indeed, that we have a standard allocation of probability $\rho_o: \mathcal{Q}_o \to \mathcal{M}$, where $\mathcal{Q}_o$ is a subalgebra of a Boolean algebra of propositions $\mathcal{Q}$. And suppose further that the allocation ρ_o on the subalgebra $\mathcal{Q}_o$ exhausts our opinions about the subject matter of the propositions in $\mathcal{Q}$. Then does ρ_o endow us with positive degrees of belief for any of the propositions in $\mathcal{Q}$ that are not in $\mathcal{Q}_o$?

It may well do so. For suppose $A \in \mathcal{Q}$ and $A \notin \mathcal{Q}_o$. Then there may be an element $A_o \in \mathcal{Q}_o$ such that $A_o \leq A$; and in such a case the probability mass $\rho_o(A_o)$, being committed to A_o, will certainly be committed to A as well. In general we must commit to A all the probability masses $\rho_o(A_o)$ for all the $A_o \in \mathcal{Q}_o$ that are subelements of A. So altogether we must commit the probability mass $\vee\{\rho_o(A_o) \mid A_o \in \mathcal{Q}_o;\ A_o \leq A\}$ to A. So the possession of the allocation $\rho_o: \mathcal{Q}_o \to \mathcal{M}$ and the lack of any further opinions about $\mathcal{Q}$ would seem to leave us with an allocation

$$\rho: \mathcal{Q} \to \mathcal{M}: A \mapsto \vee\{\rho_o(A_o) \mid A_o \in \mathcal{Q}_o;\ A_o \leq A\} \tag{1}$$

on $\mathcal{Q}$. But it this an allocation?

<u>Theorem</u>. Suppose $\mathcal{Q}_o$ is a subalgebra of a Boolean algebra $\mathcal{Q}$ and $\rho_o: \mathcal{Q}_o \to \mathcal{M}$ is a standard allocation of probability. Then the mapping $\rho: \mathcal{Q} \to \mathcal{M}$ given by (1) is a standard allocation on $\mathcal{Q}$. Furthermore, $\rho \mid \mathcal{Q}_o = \rho_o$. And the belief functions Bel_o and Bel given by ρ_o and ρ respectively are related by the formula

$$Bel(A) = \sup_{\substack{A_o \in \mathcal{Q}_o, \\ A_o \leq A}} Bel_o(A_o) , \tag{2}$$

while the upper probability functions P^*_o and P^* are related by

$$P^*(A) = \inf_{\substack{A_o \in \mathcal{A}_o, \\ A \leq A_o}} P^*_o(A_o) \,. \tag{3}$$

<u>Proof</u>: The existence of the probability masses $\vee\{\rho_o(A_o) \mid A_o \in \mathcal{A}_o; A_o \leq A\}$ depends, of course, on the fact that $(\mathcal{M}, \mu)$ is a probability algebra, so that $\mathcal{M}$ is complete. If $A \in \mathcal{A}_o$, it is evident that $\rho(A) = \rho_o(A_o)$; hence $\rho|\mathcal{A}_o = \rho_o$. In particular, $\rho(\Lambda_{\mathcal{A}}) = \Lambda_{\mathcal{M}}$, and $\rho(V_{\mathcal{A}}) = V_{\mathcal{M}}$.

Furthermore, for all pairs $A_1, A_2 \in \mathcal{A}$,

$$\begin{aligned}
\rho(A_1) \wedge \rho(A_2) &= [\vee\{\rho_o(B_1) \mid B_1 \in \mathcal{A}_o; B_1 \leq A_1\}] \wedge [\vee\{\rho_o(B_2) \mid B_2 \in \mathcal{A}_o; B_2 \leq A_2\}] \\
&= \vee\{\rho_o(B_1) \wedge \rho_o(B_2) \mid B_1, B_2 \in \mathcal{A}_o; B_1 \leq A_1; B_2 \leq A_2\} \\
&= \vee\{\rho_o(B_1 \wedge B_2) \mid B_1, B_2 \in \mathcal{A}_o; B_1 \leq A_1, B_2 \leq A_2\} \\
&= \vee\{\rho(B) \mid B \in \mathcal{A}_o; B \leq A_1 \wedge A_2\} = \rho(A_1 \wedge A_2).
\end{aligned}$$

Hence ρ is an allocation. Since $(\mathcal{M}, \mu)$ is a probability algebra, ρ is standard. Finally, notice that for a given $A \in \mathcal{A}$, $\{\rho_o(A_o) \mid A_o \in \mathcal{A}, A_o \leq A\}$ is an upward net in $\mathcal{M}$. Hence

$$\begin{aligned}
Bel(A) = \mu(\rho(A)) &= \mu(\vee\{\rho_o(A_o) \mid A_o \in \mathcal{A}_o; A_o \leq A\}) \\
&= \sup\{\mu(\rho_o(A_o)) \mid A_o \in \mathcal{A}_o; A_o \leq A\} \\
&= \sup\{Bel_o(A_o) \mid A_o \in \mathcal{A}_o, A_o \leq A\}.
\end{aligned}$$

And

$$P^*(A) = 1 - \mathrm{Bel}(\bar{A}) = 1 - \sup\{\mathrm{Bel}_o(A_o) \mid A_o \in \mathcal{A}_o,\ A_o \leq \bar{A}\}$$

$$= \inf\{1 - \mathrm{Bel}_o(A_o) \mid A_o \in \mathcal{A}_o,\ A_o \leq \bar{A}_o\}$$

$$= \sup\{P^*_o(A_o) \mid A_o \in \mathcal{A}_o,\ A \leq A_o\}.$$ ▨

I will call ρ, Bel and P* the natural extensions of ρ_o, Bel_o and P^*_o, respectively. It should be borne in mind that in general one's belief function on a Boolean algebra will not be the natural extension of its restriction to a given subalgebra. But it seems fair to characterize the cases where it is by saying that in those cases the restriction to the subalgebra "exhausts our opinions about the subject matter of the larger algebra." More concisely, I will say that an allocation $\rho: \mathcal{A} \to \mathcal{M}$ is supported by the subalgebra $\mathcal{A}_o$ of $\mathcal{A}$ whenever ρ is the natural extension of $\rho \mid \mathcal{A}_o$.

We have already seen one simple example where we wanted to adopt the natural extension of an allocation on a subalgebra -- namely, the Senate example in section 2 of Chapter 1. In that example, we obtained a belief function on a Boolean algebra corresponding to the field of all subsets of the set of twenty-two Senators. But in fact, that belief function was derived from a belief function (which happened to be a probability function) on the subalgebra corresponding to the field of all subsets of the set of eleven States. It is easily seen that the belief function we obtained on the larger Boolean algebra is the natural extension of the belief function on the subalgebra.

Let me give another example. Suppose we have a belief function concerning the possible values of an unknown quantity $\underset{\sim}{X}$ -- i.e., a belief function Bel_o: $\mathcal{P}(\mathcal{S}_1) \to [0, 1]$, where $\mathcal{S}_1$ is the set of all possible values of the quantity $\underset{\sim}{X}$ and $Bel_o(A)$ is our degree of belief that the true value is in A. And suppose we have no opinions whatsoever about the value of a second unknown quantity $\underset{\sim}{Y}$, except the knowledge that it is in a set $\mathcal{S}_2$. And suppose we would like to define a belief function Bel: $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to [0, 1]$ which would express our opinions about the values of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ simultaneously: we would like Bel(A) to be our degree of belief that the pair (x, y) is in A, where x is the true value of $\underset{\sim}{X}$ and y is the true value of $\underset{\sim}{Y}$. What should we do?

Well, $\mathcal{P}(\mathcal{S}_1)$ is naturally isomorphic to a subalgebra of $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$. Figure 1 gives the familiar geometric picture: the horizontal axis corresponds to $\mathcal{S}_1$, the vertical axis to $\mathcal{S}_2$, the whole plane to $\mathcal{S}_1 \times \mathcal{S}_2$, and a subset A of $\mathcal{S}_1$ corresponds to a vertical "cylinder set" based on the subset A of the horizontal axis.

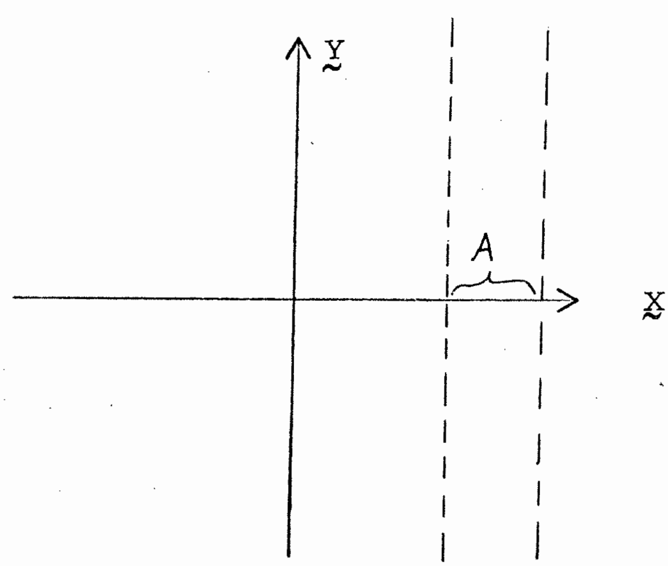

Figure 1

In symbols, the isomorphism $i: \mathcal{P}(\mathcal{S}_1) \xrightarrow{\text{into}} \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$ is given by $i(A) = \{(x, y) \mid x \in A,\ y \in \mathcal{S}_2\} = A \times \mathcal{S}_2$. So we should obviously adopt as our belief function on $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$ the natural extension of $Bel_o \circ i^{-1}$ on the subalgebra $i(\mathcal{P}(\mathcal{S}_1))$ of $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$. This will result in the belief function $Bel: \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to [0, 1]$ defined by

$$\begin{aligned} Bel(A) &= \sup \{Bel_o \circ i^{-1}(A_o) \mid A_o \in i(\mathcal{P}(\mathcal{S}_1)),\ A_o \subset A\} \\ &= \sup \{Bel_o(A_o) \mid A_o \subset \mathcal{S}_1,\ i(A_o) \subset A\} \\ &= \sup \{Bel_o(A_o) \mid A_o \subset \mathcal{S}_1,\ A_o \times \mathcal{S}_2 \subset A\} \\ &= Bel_o(\{x \mid \{x\} \times \mathcal{S}_2 \subset A\}). \end{aligned}$$

In other words, A is awarded the degree of belief of the largest vertical cylinder set that is contained in A.

2. Restricted Allocations

In the preceding section, we saw how to obtain an allocation or belief function on a Boolean algebra of propositions $\mathcal{A}$ starting with an allocation or belief function on a subalgebra $\mathcal{A}_o$. Actually, the same sort of extension can be carried out even when the original allocation is on a subset of $\mathcal{A}$ which falls short of being a subalgebra by failing to include negations of some of its elements or disjunctions of some pairs of its elements.

Of course, our definitions for the notions of an allocation and a belief function apply only to a Boolean algebra, but they do not involve negations or disjunctions in any essential way and hence can be trivially generalized. This is done in the following definitions.

Definition. I will call a subset $\mathcal{L}$ of a Boolean algebra $\mathcal{A}$ a subtrellis of $\mathcal{A}$ if

(i) $\Lambda_{\mathcal{A}} \in \mathcal{L}$,

(ii) $V_{\mathcal{A}} \in \mathcal{L}$,

and (iii) $A_1 \wedge A_2 \in \mathcal{L}$ whenever $A_1, A_2 \in \mathcal{L}$. (This terminology is not standard.)

Definition. Suppose $\mathcal{L}$ is a subtrellis of a Boolean algebra $\mathcal{A}$. Then a function $\mathrm{Bel}: \mathcal{L} \to [0, 1]$ is a restricted belief function if

(i) $\mathrm{Bel}(\Lambda_{\mathcal{A}}) = 0$,

(ii) $\mathrm{Bel}(V_{\mathcal{A}}) = 1$,

and (iii) $\mathrm{Bel}(A) \geq \Sigma\, \mathrm{Bel}(A_i) - \Sigma\, \mathrm{Bel}(A_i \wedge A_j) + - \ldots + (-1)^{n+1}$
$\mathrm{Bel}(A_1 \wedge \ldots \wedge A_n)$

for all collections, $A, A_1, \ldots, A_n$ of elements of $\mathcal{L}$ such that $A_i \leq A$ for $i = 1, \ldots, n$.

Definition. Suppose $\mathcal{L}$ is a subtrellis of a Boolean algebra $\mathcal{A}$ and $(\mathcal{M}, \mu)$ is a measure algebra. Then a mapping $\rho: \mathcal{L} \to \mathcal{M}$ is a restricted allocation of probability if

(i) $\rho(\Lambda_{\mathcal{A}}) = \Lambda_{\mathcal{M}}$

(ii) $\rho(V_{\mathcal{A}}) = V_{\mathcal{M}}$

(iii) $\rho(A_1 \wedge A_2) = \rho(A_1) \wedge \rho(A_2)$ whenever $A_1, A_2 \in \mathcal{L}$. If $\mathcal{M}$ is a probability algebra, then ρ is called standard.

Interestingly enough, our theory for allocations and belief functions remains largely valid for the restricted variety. In particular, if $\rho: \mathcal{L} \to \mathcal{M}$ is a restricted allocation and μ is the measure on $\mathcal{M}$, then $Bel = \mu \circ \rho$ will be a restricted belief function. And any restricted belief function can be represented in this way, where $(\mathcal{M}, \mu)$ is a probability algebra. These facts can be verified by noting that the proofs of Chapter 2 remain valid almost word for word for the restricted case.

It might seem desirable to cast our whole theory in a more general form by admitting restricted belief functions as belief functions. But such a generalization is unnecessary, precisely because a restricted allocation or a restricted belief function on a subtrellis $\mathcal{L}$ of a Boolean algebra $\mathcal{A}$ can always be naturally extended to a belief function or allocation on $\mathcal{A}$.

<u>Theorem</u>. Suppose $\mathcal{L}$ is a subtrellis of a Boolean algebra $\mathcal{A}$ and $\rho_o: \mathcal{L} \to \mathcal{M}$ is a standard restricted allocation. Then the mapping $\rho: \mathcal{A} \to \mathcal{M}$ given by

$$\rho(A) = \vee \{ \rho_o(L) \mid L \in \mathcal{L}, \; L \leq A \}$$

is a standard allocation on $\mathcal{A}$. Furthermore, $\rho \mid \mathcal{L} = \rho_o$. And if μ denotes the measure on $\mathcal{M}$, then the belief function $Bel = \mu \circ \rho$ on $\mathcal{A}$ and the restricted belief function $Bel_o = \mu \circ \rho_o$ on $\mathcal{L}$ are related by

$$\mathrm{Bel}(A) = \sup \{ \Sigma \mathrm{Bel}_o(L_i) - \Sigma \mathrm{Bel}_o(L_i \wedge L_j) +- \dots + (-1)^{n+1} \mathrm{Bel}_o (L_1 \wedge \dots \wedge L_n) \mid n \geq 1; L_1, \dots, L_n \in \mathcal{L} \text{ ; and } L_i \leq A \text{ for } i = 1, \dots, n \}$$

for all $A \in \mathcal{Q}$.

<u>Proof:</u> The proof that ρ is a standard allocation and $\rho \mid \mathcal{L} = \rho_o$ is precisely the same as the proof of the analogous assertions in the preceding section.

To verify the formula for Bel(A), notice that $\{\rho_o(L_1) \vee \dots \vee \rho_o(L_n) \mid n \geq 1, L_1, \dots, L_n \in \mathcal{L} ; L_i \leq A \text{ for all } i\}$ is an upward net in $\mathcal{M}$. Hence, denoting by μ the measure on $\mathcal{M}$, we have

$$\mathrm{Bel}(A) = \mu(\rho(A)) = \mu (\vee \{ \rho_o (L) \mid L \in \mathcal{L} ; L \leq A\})$$

$$= \mu (\vee \{\rho_o (L_1) \vee \dots \vee \rho_o(L_n) \mid L_1, \dots, L_n \in \mathcal{L} ; L_i \leq A \text{ for all } i\})$$

$$= \sup \{ \mu (\rho_o(L_1) \vee \dots \vee \rho_o(L_n)) \mid L_1, \dots, L_n \in \mathcal{L} ; L_i \leq A \text{ for all } i \}$$

$$= \sup \{ \Sigma \mu(\rho_o(L_i)) - \Sigma \mu (\rho_o (L_i \wedge L_j)) +- \dots + (-1)^{n+1} \mu (\rho_o(L_1 \wedge \dots \wedge L_n)) \mid L_1, \dots, L_n \in \mathcal{L} ; L_i \leq A \text{ for all } i \}$$

$$= \sup \{ \Sigma \mathrm{Bel}_o(L_i) - \Sigma \mathrm{Bel}_o(L_i \wedge L_j) +- \dots + (-1)^{n+1} \mathrm{Bel}_o(L_1 \wedge \dots \wedge L_n) \mid L_1, \dots, L_n \in \mathcal{L} ; L_i \leq A \text{ for all } i\}.$$

Of course, I will call Bel and ρ the natural extension to $\mathcal{Q}$ of Bel_o and ρ_o, respectively.

3. The Combination of Belief Functions

In section 1 I discussed an example of extension that involved two unknown quantities $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ with sets $\mathcal{S}_1$ and $\mathcal{S}_2$ of possible values, respectively. Beginning with a belief function $\mathrm{Bel}_o: \mathcal{P}(\mathcal{S}_1) \rightarrow [0,1]$ and operating on the assumption that I had no opinions about the value of $\underset{\sim}{Y}$, I obtained a belief function $\mathrm{Bel}: \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \rightarrow [0,1]$. But of course even when I have no opinions about $\underset{\sim}{Y}$ I can still claim to have a belief function Bel_2 on $\mathcal{P}(\mathcal{S}_2)$; it will be the vacuous belief function:

$$\mathrm{Bel}_2(A) = \begin{cases} 0 & \text{if } A \neq \mathcal{S}_2 \\ 1 & \text{if } A = \mathcal{S}_2. \end{cases}$$

So instead of thinking of $\mathrm{Bel}: \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \rightarrow [0,1]$ as the result of extending Bel, we can think of it as the result of combining Bel_1 on $\mathcal{P}(\mathcal{S}_1)$ with the vacuous belief function Bel_2 on $\mathcal{P}(\mathcal{S}_2)$.

This example raises the question of whether there is a natural general rule for combining belief functions on different Boolean algebras. More precisely, when Bel_1 is a belief function on the Boolean algebra $\mathcal{A}_1$, and Bel_2 is a belief function on the Boolean algebra $\mathcal{A}_2$, is there a natural way of combining the two to obtain a belief function Bel on $\mathcal{A} = \mathcal{A}_1 \oplus \mathcal{A}_2$?

Recall that $\mathcal{A}_1$ and $\mathcal{A}_2$ are thought of as independent subalgebras of $\mathcal{A}$. So one could begin to define Bel on $\mathcal{A}$ by setting $\mathrm{Bel}(A) = \mathrm{Bel}_1(A)$ when $A \in \mathcal{A}_1$ and $\mathrm{Bel}(A) = \mathrm{Bel}_2(A)$ when $A \in \mathcal{A}_2$. But many elements of $\mathcal{A}$

are in neither $\mathcal{A}_1$ nor $\mathcal{A}_2$. For example if $A_1 \varepsilon \mathcal{A}_1$ and $A_2 \varepsilon \mathcal{A}_2$ and neither A_1 nor A_2 is the zero or the unit, then $A_1 \wedge A_2$ will be in neither $\mathcal{A}_1$ nor in $\mathcal{A}_2$.

So suppose $A_1 \varepsilon \mathcal{A}_1$, $Bel_1(A_1) = \alpha_1$, $A_2 \varepsilon \mathcal{A}_2$, and $Bel_2(A_2) = \alpha_2$. Then what degree of belief should we assign to $A_1 \wedge A_2$? Well, Bel_1 directs us to commit α_1 of our belief to A_1, and Bel_2 directs us to commit α_2 of our belief to A_2. Supposing that we have already carried out Bel_1's directions, then the natural procedure is to apply Bel_2's directions not just to our probability as a whole, but to every probability mass, including the probability mass of measure α_1 that is committed to A_1. Hence we would commit α_2 of that probability mass, or a probability mass of measure $\alpha_1 \cdot \alpha_2$, to A_2 as well and hence to $A_1 \wedge A_2$. At any rate, this would be the natural procedure if Bel_1 and Bel_2 were derived from independent sources of information.

So we have a method for determining a degree of belief for each element $A \varepsilon \mathcal{A}$ that can be represented in the form $A = A_1 \wedge A_2$, where $A_1 \varepsilon \mathcal{A}_1$ and $A_2 \varepsilon \mathcal{A}_2$: we set $Bel(A) = Bel_1(A_1) \cdot Bel_2(A_2)$. This quantity is well-defined; for if $A \neq \Lambda$, then the representation $A = A_1 \wedge A_2$ is unique; while if $A = \Lambda$, then either A_1 or A_2 is the zero and $Bel_o(A) = Bel_1(A_1) \cdot Bel_2(A_2) = 0$.

But the set $\mathcal{L} = \{A \mid A = A_1 \wedge A_2;\ A_1 \varepsilon \mathcal{A}_1;\ A_2 \varepsilon \mathcal{A}_2\}$ is a subtrellis of $\mathcal{A}$. Indeed, $\Lambda = \Lambda \wedge \Lambda$, $V = V \wedge V$, and $(A_1 \wedge A_2) \wedge (A_1' \wedge A_2') = (A_1 \wedge A_1') \wedge (A_2 \wedge A_2')$ is in $\mathcal{L}$ whenever $A_1, A_1' \varepsilon \mathcal{A}_1$ and $A_2, A_2' \varepsilon \mathcal{A}_2$. So we have a function $Bel_o: \mathcal{L} \to [0,1]: A_1 \wedge A_2 \rightsquigarrow Bel_1(A_1) \cdot Bel_2(A_2)$ on a subtrellis

$\mathcal{L}$. If this were a restricted belief function on $\mathcal{L}$ (and we have not shown that it is), then by the theory of the preceding section, we could extend it to a belief function Bel on $\mathcal{A}$ that would be given by

$$\begin{aligned}
\mathrm{Bel}(A) &= \sup \{\Sigma\, \mathrm{Bel}_o(L_i) - \Sigma\, \mathrm{Bel}_o(L_i \wedge L_j) + - \ldots + (-1)^{n+1}\, \mathrm{Bel}_o(L_1 \wedge \ldots \\
&\qquad \wedge L_n) \,|\, L_1, \ldots, L_n \in \mathcal{L};\ L_i \leq A \text{ for all } i\} \\
&= \sup \{\Sigma\, \mathrm{Bel}_1(A_i) \cdot \mathrm{Bel}_2(B_i) - \Sigma\, \mathrm{Bel}_1(A_i \wedge A_j) \cdot \mathrm{Bel}_2(B_i \wedge B_j) \\
&\qquad + - \ldots + (-1)^{n+1}\, \mathrm{Bel}_1(A_1 \wedge \ldots \wedge A_n)\, \mathrm{Bel}_2(B_1 \wedge \ldots \wedge B_n) \,| \\
&\qquad A_1, \ldots, A_n \in \mathcal{A}_1;\ B_1, \ldots, B_n \in \mathcal{A}_2;\ \text{and } A_i \wedge B_i \leq A, \\
&\qquad i = 1, \ldots, n\}
\end{aligned}$$

But how shall we show that Bel_o is a restricted belief function on $\mathcal{L}$?

The easiest way is to turn to the theory of allocations.

<u>Theorem.</u> Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of a Boolean algebra $\mathcal{A}$, and suppose $\rho_1^o: \mathcal{A}_1 \to \mathcal{M}_1$ and $\rho_2^o: \mathcal{A}_2 \to \mathcal{M}_2$ are standard allocations with belief functions $\mathrm{Bel}_1 = \mu_1 \circ \rho_1^o$ and $\mathrm{Bel}_2 = \mu_2 \circ \rho_2^o$, where μ_1 and μ_2 are the measures on $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. Let $((\mathcal{M}, \mu);\ i_1: \mathcal{M}_1 \to \mathcal{M};\ i_2: \mathcal{M}_2 \to \mathcal{M})$ be an orthogonal sum of $(\mathcal{M}_1, \mu_1)$ and $(\mathcal{M}_2, \mu_2)$. Then $\rho_1 = i_1 \circ \rho_1^o$ and $\rho_2 = i_2 \circ \rho_2^o$ will be standard allocations of $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively, into $\mathcal{M}$; $\mathrm{Bel}_1 = \mu \circ \rho_1$ and $\mathrm{Bel}_2 = \mu \circ \rho_2$. Now define $\rho: \mathcal{A} \to \mathcal{M}$ by

$$\rho(A) = \vee \{\rho_1(A_1) \wedge \rho_2(A_2) \,|\, A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \leq A\} \tag{1}$$

Then ρ is an allocation of probability. Denote Bel $= \mu \circ \rho$. Then Bel $| \mathcal{A}_1 =$ Bel$_1$, Bel $| \mathcal{A}_2 =$ Bel$_2$, and in general

$$\begin{aligned}\mathrm{Bel}(A) = \sup\{ &\Sigma\, \mathrm{Bel}_1(A_i) \cdot \mathrm{Bel}_2(B_i) - \Sigma\, \mathrm{Bel}_1(A_i \wedge A_j) \cdot \mathrm{Bel}_2(B_i \wedge B_j) \\ &+ - \ldots + (-1)^{n+1}\, \mathrm{Bel}_1(A_1 \wedge \ldots \wedge A_n)\, \mathrm{Bel}_2(B_1 \wedge \ldots \wedge B_n) \,\big|\, \\ &n \geq 1;\ A_1, \ldots, A_n \in \mathcal{A}_1;\ B_1, \ldots, B_n \in \mathcal{A}_2;\ \text{and } A_i \wedge B_i \leq A, \\ &i = 1, \ldots, n\}.\end{aligned} \tag{2}$$

<u>Proof</u>: Let $\mathcal{L}$ be the subtrellis of all elements of $\mathcal{A}$ of the form $A_1 \wedge A_2$, where $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. Define $\rho_o : \mathcal{L} \to \mathcal{M}$ by $\rho_o(A) = \rho_1(A_1) \wedge \rho_2(A_2)$ whenever $A = A_1 \wedge A_2$, with $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. Since the representation $A = A_1 \wedge A_2$ is unique when $A \neq \Lambda$, ρ_o is well-defined. It is easily verified that ρ_o is a restricted allocation, and obviously $\rho_o | \mathcal{A}_1 = \rho_1$ and $\rho_o | \mathcal{A}_2 = \rho_2$. By the theorem in section 2, the formula (1) defines the natural extension of ρ_o to $\mathcal{A}$, and Bel $= \mu \circ \rho$ is given by (2). And since $\mathcal{A}_1$ and $\mathcal{A}_2$ are subsets of $\mathcal{L}$, Bel $| \mathcal{A}_i = \mu \circ \rho | \mathcal{A}_i = \mu \circ \rho_o | \mathcal{A}_i = \mu \circ \rho_i =$ Bel$_i$ for $i = 1, 2$. ▨

From formula (2) it is evident that Bel does not depend on the choice of ρ_1 and ρ_2 or on the choice of the orthogonal sum $(\mathcal{M}, \mu)$. Hence

I will call Bel the <u>orthogonal sum</u> of Bel_1 and Bel_2 on $\mathcal{A}$, and sometimes I will denote it as $Bel_1 \oplus Bel_2$. Notice that $\mathcal{A}_1$ and $\mathcal{A}_2$ can be independent subalgebras of more than one Boolean algebra; hence it may be necessary to specify the algebra $\mathcal{A}$ when speaking of the orthogonal sum of Bel_1 on $\mathcal{A}_1$ and Bel_2 on $\mathcal{A}_2$. But usually this will make no practical difference, for if $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of $\mathcal{A}_o$ and $\mathcal{A}_o$ is a subalgebra of $\mathcal{A}$, then the orthogonal sum of Bel_1 and Bel_2 on $\mathcal{A}$ is simply the extension of the orthogonal sum on $\mathcal{A}_o$, and both are given by (2).

In particular, given belief functions Bel_1 and Bel_2 on Boolean algebras $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively, (2) will give the orthogonal sum $Bel_1 \oplus Bel_2$ on $\mathcal{A}_1 \oplus \mathcal{A}_2$. And given belief functions Bel_1 and Bel_2 on power sets $\mathcal{P}(\mathcal{S}_1)$ and $\mathcal{P}(\mathcal{S}_2)$, respectively, (2) will give the orthogonal sum $Bel_1 \oplus Bel_2$ on the power set $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$. In this latter case, (2) becomes

$$\begin{aligned} Bel(A) = \sup \{ & \Sigma Bel_1(A_i)\, Bel_2(B_i) - \Sigma Bel_1(A_i \cap A_j)\, Bel_2(B_i \cap B_j) \\ & + - \ldots + (-1)^{n+1}\, Bel(A_1 \cap \ldots \cap A_n)\, Bel(B_1 \cap \ldots \cap B_n) \mid \\ & n \geq 1,\ A_1, \ldots, A_n \subset \mathcal{S}_1,\ B_1, \ldots, B_n \subset \mathcal{S}_2;\ A_i \times B_i \subset A, \\ & i = 1, \ldots, n \} \end{aligned} \tag{2'}$$

This brings us back to the example with which we began. In that case, Bel_2 is the vacuous belief function, and (2') becomes

$$\begin{aligned} Bel(A) = \sup \{ & \Sigma Bel_1(A_i) - \Sigma Bel_1(A_i \cap A_j) + - \ldots + (-1)^{n+1} \\ & Bel_1(A_1 \cap \ldots \cap A_n) \mid A_1, \ldots, A_n \subset \mathcal{S}_1;\ A_i \times \mathcal{S}_2 \subset A, \\ & i = 1, \ldots, n \} \end{aligned}$$

$$= \sup\{ Bel_1(A_1) \mid A_1 \subset \mathcal{S}_1,\ A_1 \times \mathcal{S}_2 \subset A\}$$

$$= Bel_1(\{ x \mid \{x\} \times \mathcal{S}_2 \subset A\}).$$

This does indeed agree with the method of extension.

Theorem. Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of $\mathcal{A}$, and Bel: $\mathcal{A} \to [0, 1]$ is the orthogonal sum of $Bel_1 : \mathcal{A}_1 \to [0, 1]$ and $Bel_2 : \mathcal{A}_2 \to [0, 1]$, and let P^*, $P_1{}^*$ and $P_2{}^*$ denote the upper probability functions corresponding to Bel, Bel_1 and Bel 2, respectively. Then for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$,

$$\text{(i)} \quad Bel\,(A_1 \wedge A_2) = Bel_1(A_1) \cdot Bel_2(A_2),$$

and

$$\text{(ii)} \quad P^*(A_1 \wedge A_2) = P_1{}^*(A_1) \cdot P_2{}^*(A_2).$$

Proof: (i) is clear from the preceding theorem, but (ii) is more difficult. Let $(\mathcal{M}, \mu)$, ρ, ρ_1, ρ_2 be as in the preceding theorem, and let ζ, ζ_1 and ζ_2 be the allowments corresponding to ρ, ρ_1 and ρ_2, respectively. Then $P^* = \mu \circ \zeta$, $P_1{}^* = \mu \circ \zeta_1$, $P_2{}^* = \mu \circ \zeta_2$, and since $\zeta_1(\mathcal{A}_1)$ and $\zeta_2(\mathcal{A}_2)$ are in orthogonal subalgebras of $\mathcal{M}$, we can establish (ii) by showing that $\zeta(A_1 \wedge A_2) = \zeta(A_1) \wedge \zeta(A_2)$ whenever $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. But in such a case,

$$\zeta(A_1 \wedge A_2) = \overline{\rho\ (\overline{A_1 \wedge A_2})}$$

$$= \overline{\vee \{ \rho_1(A) \wedge \rho_2(B) \mid A \in \mathcal{A}_1;\ B \in \mathcal{A}_2;\ A \wedge B \leq \overline{A_1 \wedge A_2}\}}$$

$$= \wedge \{ \overline{\rho_1(A) \wedge \rho_2(B)} \mid A \in \mathcal{A}_1;\ B \in \mathcal{A}_2:\ A \wedge B \leq \overline{A_1 \wedge A_2}\}$$

$$= \wedge \{\overline{\rho_1(\overline{A}_1)} \vee \overline{\rho_2(\overline{B})} \mid A \in \mathcal{A}_1;\ B \in \mathcal{A}_2;\ \overline{A} \wedge \overline{B} \leq \overline{A_1 \wedge A_2}\}$$

$$= \wedge \{ \zeta_1(A) \vee \zeta_2(B) \,|\, A \in \mathcal{Q}_1;\ B \in \mathcal{Q}_2;\ A_1 \wedge A_2 \leq A \vee B \}.$$

But notice that $\zeta_1(A_1) \vee \zeta_2(\Lambda) = \zeta_1(A_1)$ and $\zeta_1(\Lambda) \vee \zeta_2(A_2) = \zeta_2(A_2)$ are in this last meet. And whenever $A \in \mathcal{Q}_1$, $B \in \mathcal{Q}_2$ and $A_1 \wedge A_2 \leq A \vee B$, we know (by the second theorem of Chapter 3, section 9) that either $A_1 \leq A$ or $A_2 \leq B$. Hence every other probability mass in the meet will contain either $\zeta_1(A_1)$ or $\zeta_2(A_2)$ and hence, in any case, $\zeta_1(A_1) \wedge \zeta_2(A_2)$. Hence the meet is equal to $\zeta_1(A_1) \wedge \zeta_2(A_2)$. ▨

4. A Combinatorial Lemma

Lemma. Suppose m and n are positive integers, $I \subset \{1, \ldots, n\}$, $J \subset \{1, \ldots, m\}$, and I and J are non-empty. Set

$$\mathcal{K} = \{K \,|\, \phi \neq K \subset \{1, \ldots, n\} \times \{1, \ldots, m\};\ I = \{i \,|\, (i,j) \in K \text{ for some } j\};\ J = \{j \,|\, (i,j) \in K \text{ for some } i\}\ \}.$$

Then

$$\sum_{K \in \mathcal{K}} (-1)^{1 + \text{card } K} = (-1)^{\text{card } I + \text{card } J}.$$

Proof: Set card I = i and Card J = j, and denote $L = \{1, \ldots, i\} \times \{1, \ldots, j\}$, and think of L as an i x j matrix. I will call a subset A of L a covering of L if A contains at least one entry in every row and every column of L. I will call such a covering even or odd according as it contains an even or odd number of entries. I will prove the following assertion: The number of odd coverings of L is one greater than the number of even coverings if $i + j$ is even, and one less if $i + j$ is odd. In symbols: #(odd coverings) - # (even coverings) = $(-1)^{i+j}$.

The proof will be by induction on $i+j$. Since I and J are non-empty, $i + j \geq 2$; and if $i + j = 2$, the assertion is trivially true. Indeed, it is trivially true whenever $i = 1$ or $j = 1$ So suppose that $i + j = k$, that the assertion is true whenever $i + j < k$, and that $i > 1$ and $j > 1$. Let L_o be the $(i - 1) \times (j - 1)$ matrix obtained by omitting the first row and column of L. Let R and C be the subsets of L indicated in Figure 2. Then by our inductive hypothesis, our assertion holds for the $(i-1) \times (j-1)$ matrix L_o, the $i \times (j-1)$ matrix $R \cup L_o$ and the $(i-j) \times j$ matrix $C \cup L_o$.

~~Let us~~ classify the coverings of L according as they (i) intersect both R and C, (ii) intersect R but not C, (iii) intersect C but not R, or (iv) intersect neither R nor C.

Consider category (i). Some of the coverings in this category contain (1, 1) but they will remain coverings if (1, 1) is omitted. Hence

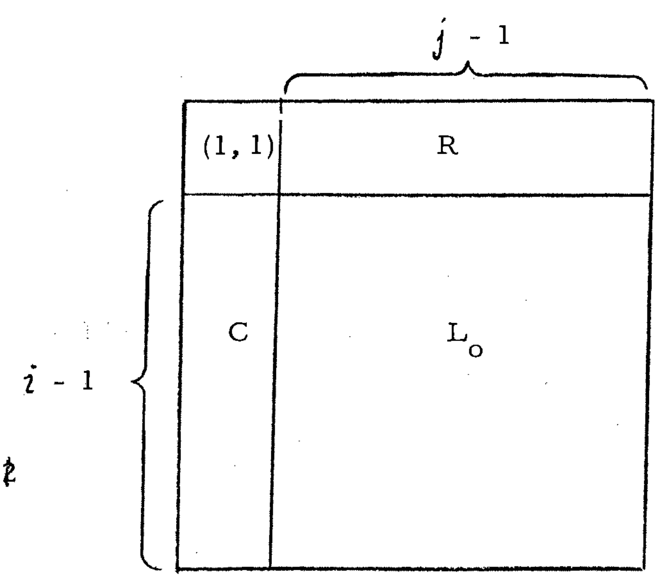

Figure 2

the coverings in this category can be arranged in pairs, the two members of which differ only in that one contains (1, 1) and the other does not. Hence there are the same number of even as odd coverings in this category.

Consider category (ii). Each covering in this category must contain (1, 1). As a matter of fact, each one is obtained from a covering of $R \cup L_o$ by adding (1, 1). Hence for this category #(odd coverings) - #(even coverings) = #(even coverings of $R \cup L_o$) - #(odd coverings of $R \cup L_o$) = $-(-1)^{i+(j-1)} = (-1)^{i+j}$.

It can be shown quite analogously for category (iii) that #(odd coverings) - #(even coverings) = $-(-1)^{i-1)+j} = (-1)^{i+j}$.

Finally, consider category (iv). Each covering in this category must contain L_o and must also be a covering of L_o. As a matter of fact, the elements of this category are obtained by taking coverings of L_o and adding (1, 1). Hence for this category #(odd coverings) - #(even coverings) = #(even coverings for L_o) - #(odd coverings for L_o) = $-(-1)^{(i-1)+(j-1)} = (-1)^{i+j-1}$.

Adding the results for all four categories, we find that ~~overall~~ #(odd coverings) - #(even coverings) = $(-1)^{i+j} + (-1)^{i+j} + (-1)^{i+j-1}$ $= (-1)^{i+j}$.

The lemma follows immediately ~~from~~ this result.

<u>Corollary.</u> Suppose $(\mathcal{M}, \mu)$ is a probability algebra, $\mathcal{E}_o$ and $\mathcal{F}_o$ are subtrellises of $\mathcal{M}$, and

$$\mu(E \wedge F) = \mu(E) \cdot \mu(F) \tag{3}$$

for all $E \epsilon \mathcal{E}_o$ and $F \epsilon \mathcal{F}_o$. Denote by $\mathcal{E}$ and $\mathcal{F}$ the subalgebras of $\mathcal{M}$ generated by $\mathcal{E}_o$ and $\mathcal{F}_o$, respectively. Then (1) holds for all $E \epsilon \mathcal{E}$ and $F \epsilon \mathcal{F}$.

Proof: Consider first elements E and F of $\mathcal{M}$ of the form

$$E = E_o \wedge \overline{E}_1 \wedge \dots \wedge \overline{E}_K = (E_o \vee E_1 \vee \dots \vee E_K) - (E_1 \vee \dots \vee E_K)$$

and

$$F = F_o \wedge \overline{F_1} \wedge \dots \wedge \overline{F}_\ell = (F_o \vee F_1 \vee \dots \vee F_\ell) - (F_1 \vee \dots \vee F_\ell),$$

where $E_o, E_1, \dots, E_K \epsilon \mathcal{E}_o$ and $F_o, F_1, \dots, F_\ell \epsilon \mathcal{F}_o$. We have

$$E \wedge F = E_o \wedge F_o \wedge \overline{E}_1 \wedge \dots \wedge \overline{E}_K \wedge \overline{F_1} \wedge \dots \wedge \overline{F}_\ell$$

$$= (E_o \wedge F_o) \wedge \overline{E_1 \vee \dots \vee E_K \vee F_1 \vee \dots \vee F_\ell}$$

$$= [(E_o \wedge F_o) \vee E_1 \vee \dots \vee E_K \vee F_1 \vee \dots \vee F_\ell] -$$

$$[(E_1 \vee \dots \vee E_K \vee F_1 \vee \dots \vee F_\ell],$$

and

$$\mu(E \wedge F) = \sum_{I \subset \{1, \dots, k\}} \sum_{J \subset \{1, \dots, \ell\}} (-1)^{\text{card } I + \text{card } J} \cdot$$

$$\mu (E_o \wedge F_o \wedge (\wedge_{i \epsilon I} E_i) \wedge (\wedge_{j \epsilon J} F_j))$$

$$= \sum_{I \subset \{1, \dots, k\}} (-1)^{\text{card } I} \mu (E_o \wedge (\wedge_{i \epsilon I} E_i))$$

$$\times \sum_{J \subset \{1, \dots, \ell\}} (-1)^{\text{card } J} \mu (F_o \wedge (\wedge_{j \epsilon J} F_j))$$

$$= \mu(E) \cdot \mu(F).$$

Now by section 7 of Chapter 3, any element $E \epsilon \mathcal{E}$ can be written in the form

$$E = E_1 \vee \ldots \vee E_m,$$

where for each i, $i = 1, \ldots, m$,

$$E_i = E_{io} \wedge \overline{E_{i1}} \wedge \ldots \wedge \overline{E_{ik_i}}$$

for some elements $E_{io}, E_{i1}, \ldots, E_{ik_i}$ of $\mathcal{E}_o$. Similarly, any element $F \varepsilon \mathcal{F}$ can be written in the form

$$F = F_1 \vee \ldots \vee F_n,$$

where for each i, $i = 1, \ldots, n$,

$$F_i = F_{io} \wedge \overline{F_{i1}} \wedge \ldots \wedge \overline{F_{i\ell_i}}$$

for some elements $F_{io}, F_{i1}, \ldots, F_{i\ell_i}$ of $\mathcal{F}_o$. If E and F are expressed in this way, then

$$E \wedge F = (E_1 \vee \ldots \vee E_m) \wedge (F_1 \vee \ldots \vee F_n)$$

$$= \bigvee_{i=1}^{m} \bigvee_{j=1}^{n} (E_i \wedge F_j).$$

And by the lemma,

$$\mu(E \wedge F) = \mu(\bigvee_{i=1}^{m} \bigvee_{j=1}^{n} (E_i \wedge E_j))$$

$$= \sum_{\substack{K \subset \{1, \ldots, m\} x \{1, \ldots, n\} \\ K \neq \phi}} (-1)^{1 + \text{card } K} \mu(\bigwedge_{(i,j) \varepsilon K} (E_i \wedge F_j))$$

$$= \sum_{\substack{I \subset \{1, \ldots, m\} \\ I \neq \phi}} (-1)^{\text{card } I} \sum_{J \subset \{1, \ldots, n\}} (-1)^{\text{card } J}$$

$$\mu((\bigwedge_{i \varepsilon I} E_i) \wedge (\bigwedge_{j \varepsilon J} F_j)).$$

But

$$\bigwedge_{i\in I} E_i = \left(\bigwedge_{i\in I} E_{io}\right) \wedge \left(\bigwedge_{i\in I} (\bar{E}_{i1} \wedge \ldots \wedge \bar{E}_{ik_i})\right),$$

where E_{io}, E_{i1}, ..., E_{ik_i} are all in $\mathcal{E}_o$ for all i; and $\bigwedge_{j\in J} F_j$ is of a similar form. Hence by the first paragraph

$\mu((\bigwedge_{i\in I} E_i) \wedge (\bigwedge_{j\in J} F_j)) = \mu(\bigwedge_{i\in I} E_i)\, \mu(\bigwedge_{j\in J} F_j)$. So

$$\mu(E \wedge F) = \sum_{\substack{I \subset \{1,\ldots,m\} \\ I \neq \emptyset}} (-1)^{1+\text{card}\, I} \mu(\bigwedge_{i\in I} E_i)$$

$$\times \sum_{\substack{J \subset \{1,\ldots,n\} \\ J \neq \emptyset}} (-1)^{1+\text{card}\, J} \mu(\bigwedge_{j\in J} F_j)$$

$$= \mu(E) \cdot \mu(F).$$

5. Orthogonality and Independence

As we have just seen, our rule of combination obeys a multiplicative rule for both Bel and P*. In this section, I will explore the implications of these two rules.

Definition. Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of a Boolean algebra $\mathcal{A}$, and suppose Bel: $\mathcal{A} \to [0,1]$ is a belief function. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal with respect to Bel if

$$\text{Bel}(A_1 \wedge A_2) = \text{Bel}(A_1) \cdot \text{Bel}(A_2) \tag{1}$$

whenever $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. And $\mathcal{A}_1$ and $\mathcal{A}_2$ are cognitively independent with respect to Bel if

$$P^*(A_1 \wedge A_2) = P^*(A_1) \cdot P^*(A_2) \tag{2}$$

whenever $A_1 \varepsilon \mathcal{A}_1$ and $A_2 \varepsilon \mathcal{A}_2$. ($P^*$ is, of course, the upper probability function corresponding to Bel.)

A justification for the term "cognitively independent" will be offered in the next chapter. The term "orthogonal," on the other hand, can be justified immediately.

Theorem. Suppose $\mathcal{A}$ is a Boolean algebra, $(\mathcal{M}, \mu)$ is a probability algebra, and $\rho: \mathcal{A} \to \mathcal{M}$ is a standard representation for the belief function Bel on $\mathcal{A}$. Then two independent subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal with respect to Bel if and only if the subalgebras of $\mathcal{M}$ generated by $\rho(\mathcal{A}_1)$ and $\rho(\mathcal{A}_2)$ are orthogonal with respect to μ.

Proof: Denote by $\mathcal{M}_1$ and $\mathcal{M}_2$ the subalgebras of $\mathcal{M}$ generated by $\rho(\mathcal{A}_1)$ and $\rho(\mathcal{A}_2)$, respectively. Clearly, if $A_1 \varepsilon \mathcal{A}_1$ and $A_2 \varepsilon \mathcal{A}_2$, then $\rho(A_1) \varepsilon \mathcal{M}_1$ and $\rho(A_2) \varepsilon \mathcal{M}_2$, so that the orthogonality of $\mathcal{M}_1$ and $\mathcal{M}_2$ will imply (1).

Suppose, on the other hand, that (1) holds for all $A_1 \varepsilon \mathcal{A}_1$ and all $A_2 \varepsilon \mathcal{A}_2$. Then since $\rho(\mathcal{A}_1)$ and $\rho(\mathcal{A}_2)$ are subtrellises, it follows by the corollary in the preceding section that

$$\mu(M_1 \wedge M_2) = \mu(M_1) \cdot \mu(M_2)$$

for all $M_1 \varepsilon \mathcal{M}_1$ and $M_2 \varepsilon \mathcal{M}_2$. Since μ is positive it follows that $\mathcal{M}_1$ and $\mathcal{M}_2$ are independent subalgebras and hence orthogonal.

▨

It is not obvious at first glance that orthogonality and cognitive independence are distinct conditions, and hence it is worthwhile to provide examples showing that each of the conditions can hold without the other holding. To this end, set $\mathcal{A} = \mathcal{P}(\mathcal{S})$, where $\mathcal{S} = \{a, b, c, d\}$ as shown in Figure 3. And set $\mathcal{A}_1 = \{\phi, \{a,b\}, \{c,d\}, \mathcal{S}\}$ and $\mathcal{A}_2 = \{\phi \{a, c\}, \{b, d\}, \mathcal{S}\}$. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of $\mathcal{A}$. Let us define two belief functions Bel_1 and Bel_2 on $\mathcal{A}$ as follows: Bel_1 is given by the basic probability numbers $\{m_A\}_{A \in \mathcal{A}}$, where

$$
\begin{aligned}
m_{\mathcal{S}\{a,b,c\}} &= 1/4, \\
m_{\{a,b\}} &= 1/4, \\
m_{\{a,c\}} &= 1/4, \\
m_{\{a\}} &= 1/4,
\end{aligned}
$$

and $m_A = 0$ for all other $A \in \mathcal{A}$. And Bel_2 is given by the basic probability numbers $\{m'_A\}_{A \in \mathcal{A}}$, where

$$
\begin{aligned}
m'_{\mathcal{S}} &= 1/4, \\
m'_{\{a,b,c\}} &= 1/4, \\
m'_{\{a\}} &= 1/2,
\end{aligned}
$$

and $m'_A = 0$ for all other $A \in \mathcal{A}$. Then it can be verified that $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal but not cognitively independent with respect to Bel_1 and cognitively independent but not orthogonal with respect to Bel_2.

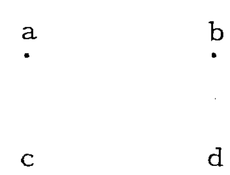

Figure 3

As we saw in section 3, when a belief function on $\mathcal{A}$ is the orthogonal sum of belief functions on independent subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$, those subalgebras are both orthogonal and cognitively independent with respect to that belief function. In fact a converse of this theorem is also true; if two independent subalgebras are both orthogonal and cognitively independent with respect to a belief function, then on the subalgebra generated by the union of the two subalgebras that belief function will agree with the orthogonal sum of its restrictions to the two subalgebras. This assertion follows from the following theorem.

Theorem. Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of a Boolean algebra $\mathcal{A}$, and suppose $\mathcal{A}$ is the subalgebra generated by $\mathcal{A}_1 \cup \mathcal{A}_2$. Suppose Bel: $\mathcal{A} \to [0,1]$ is a belief function and $\rho: \mathcal{A} \to \mathcal{M}$ is a standard representation for Bel. Then the following conditions are all equivalent:

(1) $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal and cognitively independent with respect to Bel.

(2) $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal with respect to Bel and $\rho(A \vee B) = \rho(A) \vee \rho(B)$ whenever $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$.

(3) $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal with respect to Bel and

$\rho(A) = \vee\{\rho(A_1 \wedge A_2)\ A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \leq A\}$ for all $A \in \mathcal{A}$.

(4) For all $A \in \mathcal{A}$,

$$\begin{aligned} \mathrm{Bel}(A) = \sup\{&\Sigma\, \mathrm{Bel}(A_i) \cdot \mathrm{Bel}(B_i) - \Sigma\, \mathrm{Bel}(A_i \wedge A_j) \cdot \mathrm{Bel}(B_i \wedge B_j) \\ &+- \ldots + (-1)^{n+1}\, \mathrm{Bel}(A_1 \wedge \ldots \wedge A_n)\, \mathrm{Bel}(B_1 \wedge \ldots \wedge B_n) \mid \\ &n \geq 1;\ A_1, \ldots, A_n \in \mathcal{A}_1;\ B_1, \ldots, B_m \in \mathcal{A}_2;\ \text{and } A_i \wedge B_i \leq A, \\ &i = 1, \ldots, n\}. \end{aligned}$$

Proof: (1) $\Rightarrow$ (2). Suppose $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$. Then by orthogonality and cognitive independence,

$$\begin{aligned} 1 - \mathrm{Bel}(A \vee B) = P^*(\overline{A} \wedge \overline{B}) &= P^*(\overline{A}) \cdot P^*(\overline{B}) \\ &= (1 - \mathrm{Bel}(A))\,(1 - \mathrm{Bel}(B)) \\ &= 1 - \mathrm{Bel}(A) - \mathrm{Bel}(B) + \mathrm{Bel}(A \wedge B), \end{aligned}$$

Hence

$$\mathrm{Bel}(A \vee B) = \mathrm{Bel}(A) + \mathrm{Bel}(B) - \mathrm{Bel}(A \wedge B),$$

or

$$\mu(\rho(A \vee B)) = \mu(\rho(A) \vee \rho(B)).$$

Since μ is positive, it follows that

$$\rho(A \vee B) = \rho(A) \vee \rho(B).$$

(2) $\Rightarrow$ (3). Since $\mathcal{A}$ is the subalgebra generated by $\mathcal{A}_1 \cup \mathcal{A}_2$, every element $A \in \mathcal{A}$ must be of the form

$$A = (A_1 \wedge B_1) \vee \ldots \vee (A_n \wedge B_n),$$

where the A_i are all in $\mathcal{A}_1$ and the B_i are all in $\mathcal{A}_2$. Hence, ? by (2),

$$\rho(A) = \rho(A_1 \wedge B_1) \vee \ldots \vee \rho(A_n \wedge B_n),$$

and (3) follows.

(3) $\Rightarrow$ (4). For any $A \in \mathcal{A}$,

$$\begin{aligned} \mathrm{Bel}(A) &= \mu(\rho(A)) \\ &= \mu(\vee\{\rho(A_1 \wedge A_2) \mid A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \leq A\}) \\ &= \mu(\vee\{\rho(A_1 \wedge B_1) \vee \ldots \vee \rho(A_n \wedge B_n) \mid A_i \in \mathcal{A}_1,\ B_i \in \mathcal{A}_2 \\ &\qquad \text{and } A_i \wedge B_i \leq A \text{ for all } i\}) \end{aligned}$$

$$= \sup \{ \mu (\rho(A_1 \wedge B_1) \vee \dots \vee \rho(A_n \wedge B_n)) \mid A_i \in \mathcal{A}_1, \ B_i \in \mathcal{A}_2$$
$$\text{and } A_i \wedge B_i \leq A \text{ for all } i \}$$

$$= \sup \{ \Sigma \mathrm{Bel}(A_i) \cdot \mathrm{Bel}(B_i) - \Sigma \mathrm{Bel}(A_i \wedge A_j) \cdot \mathrm{Bel}(B_i \wedge B_j)$$
$$+- \dots + (-1)^{n+1} \mathrm{Bel}(A_1 \wedge \dots \wedge A_n) \mathrm{Bel}(B_1 \wedge \dots \wedge B_n) \mid$$
$$n \geq 1; A_1, \dots, A_n \in \mathcal{A}_1; B_1, \dots, B_n \in \mathcal{A}_2;$$
$$A_i \wedge B_i \leq A, \ i = 1, \dots, n \}.$$

(4) $\Longrightarrow$ (1). This is merely a restatement of the last theorem of section 3. ▨

Finally, it is useful to note that the formulae in (2) and (3) can also be stated in terms of the allowment ζ. In terms of ζ, (2) becomes

(2') $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal with respect to Bel and $\zeta(A \wedge B) = \zeta(A) \wedge \zeta(B)$ whenever $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$;

and (3) becomes

(3') $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal with respect to Bel and

$$\zeta(A) = \wedge \{ \zeta(A_1 \vee A_2) \mid A_1 \in \mathcal{A}_1; A_2 \in \mathcal{A}_2; A_1 \vee A_2 \geq A \}.$$

6. The Finite Case

Recall that a belief function Bel on a finite Boolean algebra $\mathcal{A}$ is completely determined by the basic probability numbers m_A for $A \in \mathcal{A}$. These numbers are non-negative, $m_{\Lambda} = 0$, and Bel is given by

$$\mathrm{Bel}(A) = \sum_{A' \leq A} m_{A'}.$$

Intuitively, the basic probability number m_A measures the total probability mass that is constrained to A but not to any proper sub-element of A. In other words, if $\rho: \mathcal{Q} \to \mathcal{M}$ is an allocation representing Bel, then

$$m_A = \mu(\rho(A) - \vee\{\rho(A') \mid A' < A\}),$$

where μ is the measure on $\mathcal{M}$. It is worth noting how these basic probability numbers behave under combination.

Theorem. Suppose $\rho: \mathcal{Q} \to \mathcal{M}$ is a standard allocation on the finite Boolean algebra $\mathcal{Q}$, and suppose the independent subalgebras $\mathcal{Q}_1$ and $\mathcal{Q}_2$ of $\mathcal{Q}$ are orthogonal and cognitively independent with respect to ρ. Denote by $\{m_A\}_{A \in \mathcal{Q}}$ the basic probability numbers for ρ, by $\{n_{A_1}\}_{A_1 \in \mathcal{Q}_1}$ the basic probability numbers for $\rho \mid \mathcal{Q}_1$ and by $\{p_{A_2}\}_{A_2 \in \mathcal{Q}_2}$ the basic probability numbers for $\rho \mid \mathcal{Q}_2$. Then

$$m_A = \begin{cases} n_{A_1} \cdot p_{A_2} & \text{whenever } A = A_1 \wedge A_2 \text{ with } A_1 \in \mathcal{Q}_1 \text{ and } A_2 \in \mathcal{Q}_2 \\ 0 & \text{if } A \neq A_1 \wedge A_2 \text{ for any } A_1 \in \mathcal{Q}_1 \text{ and } A_2 \in \mathcal{Q}_2 . \end{cases}$$

Proof: First consider the case where $A \neq A_1 \wedge A_2$ for any $A_1 \in \mathcal{Q}_1$ and $A_2 \in \mathcal{Q}_2$. In that case, $A_1 \wedge A_2 < A$ whenever $A_1 \in \mathcal{Q}_1$, $A_2 \in \mathcal{Q}_2$ and $A_1 \wedge A_2 \leq A$. Hence

$$\begin{aligned}\rho(A) &= \vee\{\rho(A_1) \wedge \rho(A_2) \mid A_1 \in \mathcal{Q}_1;\ A_2 \in \mathcal{Q}_2;\ A_1 \wedge A_2 \leq A\} \\ &= \vee\{\rho(A_1) \wedge \rho(A_2) \mid A_1 \in \mathcal{Q}_1;\ A_2 \in \mathcal{Q}_2;\ A_1 \wedge A_2 < A\} \\ &= \vee\{\rho(A_1) \wedge \rho(A_2) \mid A_1 \in \mathcal{Q}_1;\ A_2 \in \mathcal{Q}_2;\ A' \in \mathcal{Q};\ A_1 \wedge A_2 \leq A' < A\}\end{aligned}$$

$$= \vee \{ \vee \{ \rho(A_1) \wedge \rho(A_2) \mid A_1 \in \mathcal{Q}_1;\ A_2 \in \mathcal{Q}_2;\ A_1 \wedge A_2 \leq A' \} \mid A' < A \}$$

$$= \vee \{ \rho(A') \mid A' < A \},$$

so

$$m_A = \mu(\rho(A) - \vee \{ \rho(A') \mid A' < A \}) = \mu(\Lambda) = 0.$$

Now consider the case where $A = A_1 \wedge A_2$ with $A_1 \in \mathcal{Q}_1$ and $A_2 \in \mathcal{Q}_2$. In that case,

$$(\rho(A_1) - \vee \{ \rho(A_1') \mid A_1' \in \mathcal{Q}_1;\ A_1' < A_1 \})$$
$$\wedge (\rho(A_2) - \vee \{ \rho(A_2') \mid A_2' \in \mathcal{Q}_2;\ A_2' < A_2 \})$$

$$= \rho(A_1) \wedge \rho(A_2) - \vee \{ \rho(A_1') \mid A_1' \in \mathcal{Q}_1;\ A_1' < A_1 \}$$
$$- \vee \{ \rho(A_2') \mid A_2' \in \mathcal{Q}_2;\ A_2' < A_2 \}$$

$$= \rho(A_1) \wedge \rho(A_2) - \vee \{ \rho(A_1') \vee \rho(A_2') \mid A_1' \in \mathcal{Q}_1;\ A_2' \in \mathcal{Q}_2;$$
$$A_1' < A_1;\ A_2' < A_2 \}$$

$$= \rho(A_1) \wedge \rho(A_2) - \vee \{ [\rho(A_1') \vee \rho(A_2')] \wedge \rho(A_1) \wedge \rho(A_2) \mid$$
$$A_1' \in \mathcal{Q}_1;\ A_2' \in \mathcal{Q}_2;\ A_1' < A_1;\ A_2' < A_2 \}$$

$$= \rho(A_1) \wedge \rho(A_2) - \vee \{ [\rho(A_1') \wedge \rho(A_2)] \vee [\rho(A_1) \wedge \rho(A_2')] \mid$$
$$A_1' \in \mathcal{Q}_1;\ A_2' \in \mathcal{Q}_2;\ A_1' < A_1;\ A_2' < A_2 \}$$

$$= \rho(A_1) \wedge \rho(A_2) - \vee \{ \rho(A_1') \wedge \rho(A_2') \mid A_1' \in \mathcal{Q}_1;\ A_2' \in \mathcal{Q}_2;$$
$$A_1' \leq A_1;\ A_2' \leq A_2;\ \text{either } A_1' < A_1 \text{ or } A_2' < A_2 \}$$

$$= \rho(A_1 \wedge A_2) - \vee\{\rho(A_1' \wedge A_2') \mid A_1' \in \mathcal{A}_1;\ A_2' \in \mathcal{A}_2;\ A_1' \leq A_1;\ A_2' \leq A_2;\ \text{either } A_1' < A_1 \text{ or } A_2' < A_2\}$$

$$= \rho(A) - \vee\{\rho(A') \mid A' < A\}$$

The last few equalities depend on the thorem of Chapter 3, section 9.

Since $\rho(\mathcal{A}_1)$ and $\rho(\mathcal{A}_2)$ are in orthogonal subalgebras of $\mathcal{M}$, the measure of $\rho(A) - \vee\{\rho(A') \mid A' < A\}$ must equal the product of the measures of

$$\rho(A_1) - \vee\{\rho(A_1') \mid A_1' \in \mathcal{A}_1;\ A_1' < A_1\}$$

and

$$\rho(A_2) - \vee\{\rho(A_2') \mid A_2' \in \mathcal{A}_2,\ A_2' < A_2\}.$$

In other words, $m_A = n_{A_1} \cdot p_{A_2}$. ▨

7. The Condensable Case

In this section, we will see how the orthogonal sum of two condensable belief functions can be described in terms of the commonality numbers.

When we are dealing with two condensable belief functions, say one on $\mathcal{P}(\mathcal{S}_1)$ and one on $\mathcal{P}(\mathcal{S}_2)$, it is most natural to consider their orthogonal sum on $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$. This orthogonal sum will itself be condensable, as we see from the following theorem.

Theorem. Suppose Bel: $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to [0,1]$ is a belief function, $\mathcal{P}(\mathcal{S}_1)$ and $\mathcal{P}(\mathcal{S}_2)$ are orthogonal and cognitively independent with respect to Bel, and Bel$|\mathcal{P}(\mathcal{S}_1)$ and Bel$|\mathcal{P}(\mathcal{S}_2)$ are condensable. Then Bel is condensable.

Proof: Let $\zeta : \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to \mathcal{M}$ be a standard allowment for Bel, and recall that ζ is condensable if and only if

$$\zeta(A) = \bigvee_{s \in A} \zeta(\{s\})$$

for all $A \subset \mathcal{S}_1 \times \mathcal{S}_2$. Now by orthogonality and cognitive independence

$$\zeta(A) = \wedge \{ \zeta(A_1 \times \mathcal{S}_2) \vee \zeta(\mathcal{S}_1 \times A_2) \mid A_1 \subset \mathcal{S}_1;\ A_2 \subset \mathcal{S}_2; \\ (A_1 \times \mathcal{S}_2) \cup (\mathcal{S}_1 \times A_2) \supset A\}$$

for all $A \subset \mathcal{S}_1 \times \mathcal{S}_2$. Since the restrictions to $\mathcal{P}(\mathcal{S}_1)$ and $\mathcal{P}(\mathcal{S}_2)$ are condensable, this becomes

$$\zeta(A) = \wedge \{ (\bigvee_{s_1 \in A_1} \zeta(\{s_1\} \times \mathcal{S}_2)) \vee (\bigvee_{s_2 \in A_2} \zeta(\mathcal{S}_1 \times \{s_2\})) \mid \\ A_1 \subset \mathcal{S}_1;\ A_2 \subset \mathcal{S}_2;\ (A_1 \times \mathcal{S}_2) \cup (\mathcal{S}_1 \times A_2) \supset A\}.$$

$$= \vee \{ \zeta(\{s_1\} \times \mathcal{S}_2) \vee \zeta(\mathcal{S}_1 \times \{s_2\}) \mid (s_1, s_2) \in A \}$$

$$= \bigvee_{(s_1, s_2) \in A} (\wedge \{ \zeta(A_1 \times \mathcal{S}_2) \vee \zeta(\mathcal{S}_1 \times A_2) \mid A_1 \subset \mathcal{S}_1; \\ A_2 \subset \mathcal{S}_2;\ (A_1 \times \mathcal{S}_2) \cup (\mathcal{S}_1 \times A_2) \supset \{(s_1, s_2)\}\})$$

$$= \bigvee_{(s_1, s_2) \in A} \zeta(\{(s_1, s_2)\}).$$

And furthermore, the commonality numbers for Bel are obtained from those for Bel $|\mathcal{P}(\mathcal{S}_1)$ and Bel $|\mathcal{P}(\mathcal{S}_2)$ by a simple multiplicative rule.

Theorem. Suppose Bel: $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$ is condensable and $\mathcal{P}(\mathcal{S})_1$ and $\mathcal{P}(\mathcal{S}_2)$ are orthogonal and cognitively independent with respect to Bel. Let

$$Q_1 \colon \mathcal{T}(\mathcal{S}_1) \to [0,1],$$

$$Q_2 \colon \mathcal{T}(\mathcal{S}_2) \to [0,1],$$

and $\quad Q \colon \mathcal{T}(\mathcal{S}_1 \times \mathcal{S}_2) \to [0,1]$

be the commonality functions for Bel $|\mathcal{P}(\mathcal{S}_1)$, Bel $|\mathcal{P}(\mathcal{S}_2)$ and Bel, respectively. Then

$$Q(\{(a_1, b_1), \ldots, (a_n, b_n)\}) = Q_1(\{a_1, \ldots, a_n\} Q_2(\{b_1, \ldots, b_n\})$$

for all $\{(a_1, b_1), \ldots, (a_n, b_n)\} \subset \mathcal{S}_1 \times \mathcal{S}_2$.

<u>Proof:</u> Letting ς be the allowment and μ the measure on the probability algebra, we have

$$Q(\{(a_1, b_1), \ldots, (a_n, b_n)\}) = \mu(\varsigma(\{(a_1, b_1)\}) \wedge \ldots \wedge \varsigma(\{(a_n, b_n)\})),$$

$$Q_1(\{a_1, \ldots, a_n\}) = \mu(\varsigma(\{a_1\} \times \mathcal{S}_2) \wedge \ldots \wedge \varsigma(\{a_n\} \times \mathcal{S}_2)),$$

and $\quad Q_2(\{b_1, \ldots, b_n\}) = \mu(\varsigma(\mathcal{S}_1 \times \{b_1\}) \wedge \ldots \wedge \varsigma(\mathcal{S}_1 \times \{b_n\})).$

But by (2') from section 5, we know that

$$\varsigma(\{(a_i, b_i)\}) = \varsigma(\{a_i\} \times \mathcal{S}_2) \wedge \varsigma(\mathcal{S}_1 \times \{b_i\})$$

for all i. Hence

$$\varsigma(\{(a_1, b_1)\} \wedge \ldots \wedge \varsigma(\{(a_n\ b_n)\})$$

$$= (\varsigma(\{a_1\} \times \mathcal{S}_2) \wedge \ldots \wedge \varsigma(\{a_n\} \times \mathcal{S}_2)) \wedge (\varsigma(\mathcal{S}_1 \times \{b_1\})$$

$$\wedge \ldots \wedge \varsigma(\mathcal{S}_1 \times \{b_n\}));$$

and the theorem follows by orthogonality. $\blacksquare$

8. An Example of Combination

In this section I will illustrate the rule of combination with a simple example.

Suppose Mr. and Mrs. Jones are discussing over their breakfast coffee whether they should attend a ballet in the evening. Mr. Jones has no opinions about how enjoyable the ballet may prove to be, yet has opinions about whether it will rain, while Mrs. Jones has no inkling as to whether it will rain yet has definite ideas about the quality of the ballet. Assuming that they trust each other's judgments in their respective areas of competency, how might Mr. and Mrs. Jones combine their opinions in order to obtain, as it were, a joint opinion about the possibility of attending an enjoyable ballet without getting wet?

Let us be more concrete. Suppose Mr. Jones has a belief function Bel_1 on $\mathcal{P}(\mathcal{S}_1)$, where $\mathcal{S}_1 = \{\text{rain, no rain}\}$, and Mrs. Jones has a belief function Bel_2 on $\mathcal{P}(\mathcal{S}_2)$, where $\mathcal{S}_2 = \{\text{enjoyable ballet, unenjoyable ballet}\}$. And suppose Bel_1 and Bel_2 are given by

$Bel_1(\emptyset) = 0$	$Bel_2(\emptyset) = 0$
$Bel_1(\{\text{rain}\}) = 1/2$	$Bel_2(\{\text{enjoyable ballet}\} = 1/2$
$Bel_1(\{\text{no rain}\}) = 0$	$Bel_2(\{\text{unenjoyable ballet}\} = 1/3$
$Bel_1(\mathcal{S}_1) = 1$	$Bel_2(\mathcal{S}_2) = 1.$

These two belief functions can also be described by saying that Bel_1 is given by the basic probability numbers $\{n_A\}_{A \subset \mathcal{S}_1}$ and Bel_2 is given by the basic probability numbers $\{p_A\}_{A \subset \mathcal{S}_2}$, where

$$n_{\emptyset} = 0 \qquad p_{\emptyset} = 0$$

$$n_{\{\text{rain}\}} = 1/2 \qquad p_{\{\text{enjoyable ballet}\}} = 1/2$$

$$n_{\{\text{no rain}\}} = 0 \qquad p_{\{\text{unenjoyable ballet}\}} = 1/3$$

$$n_{\mathcal{S}_1} = 1/2 \qquad p_{\mathcal{S}_2} = 1/6.$$

In other words, Mr. Jones puts half of his probability on the occurrence of rain and does not commit the other half, while Mrs. Jones puts half of her probability on an enjoyable ballet and a third of it on an unenjoyable one. If we represent each person's probability by a mass that is uniformly distributed over a line segment, then we can depict this situation as in Figure 4.

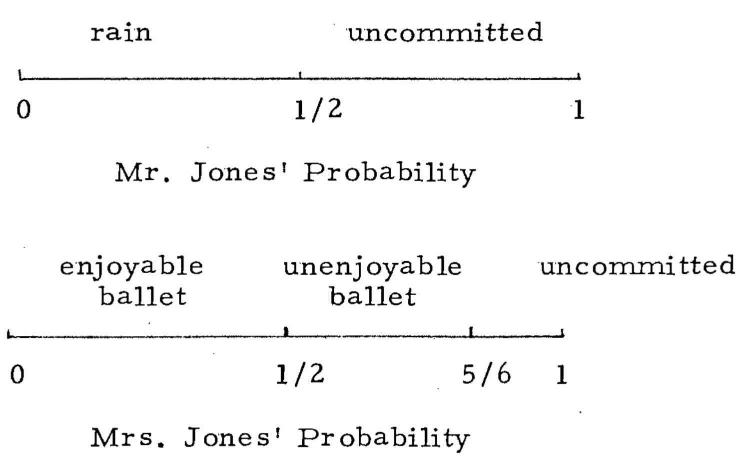

Figure 4

We require a combined belief function Bel on $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$; and in particular we require a degree of belief and an upper probability for the subset {no rain} x {enjoyable ballet} of $\mathcal{S}_1 \times \mathcal{S}_2$.

Let us consider the matter from Mrs. Jones' point of view. Her belief function Bel_2 can be described by the three basic probability masses shown in Figure 4. Now she is confronted with Mr. Jones' opinions about the weather and decides to adopt them as her own. What does this mean? Well, the message from Mr. Jones can be stated simply: Put half your probability on rain. The natural thing for Mrs. Jones to do is to carry out this recommendation for each of her three basic probability masses: she should commit half of each of them to rain.

The result can be depicted geometrically if we use a square instead of a line segment to represent Mrs. Jones' probability. In the first panel of Figure 5, Mrs. Jones' three basic probability masses are depicted, each labelled with its "region of mobility". The second panel shows the situation after she has committed half of each of her probability masses to rain but left the other halves uncommitted between rain and no rain.

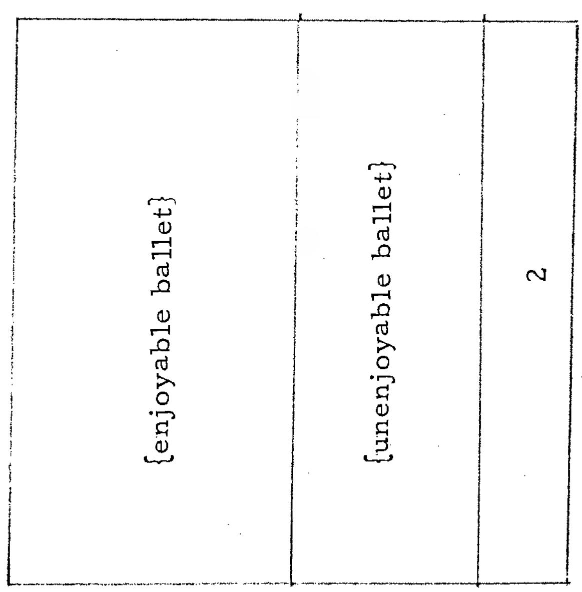

Figure 5a. Before

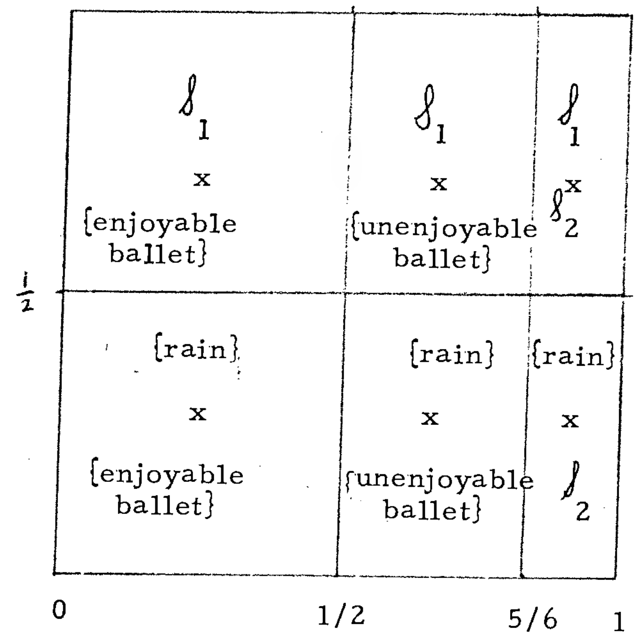

Figure 5b. After

So we obtain six basic probability masses, with the following corresponding basic probability numbers:

$$m_{\{\text{rain}\} \times \{\text{enjoyable ballet}\}} = 1/4$$

$$m_{\{\text{rain}\} \times \{\text{unenjoyable ballet}\}} = 1/6$$

$$m_{\{\text{rain}\} \times \mathcal{S}_2} = 1/12$$

$$m_{\mathcal{S}_1 \times \{\text{enjoyable ballet}\}} = 1/4$$

$$m_{\mathcal{S}_1 \times \{\text{unenjoyable ballet}\}} = 1/6$$

$$m_{\mathcal{S}_1 \times \mathcal{S}_2} = 1/12.$$

The basic probability numbers m_A for other $A \subset \mathcal{S}_1 \times \mathcal{S}_2$ are, of course, zero.

The belief function Bel on $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$ can be easily computed from this table of basic probability numbers. For example, we find that

$$\text{Bel}(\{\text{no rain}\} \times \{\text{enjoyable ballet}\}) = 0$$

and

$$P^*(\{\text{no rain}\} \times \{\text{enjoyable ballet}\}) = 1/3.$$

CHAPTER 7. DEMPSTER'S RULES OF CONDITIONING AND COMBINATION

In this chapter I adduce Demster's rules for modifying a belief function on the basis of new evidence or opinion. Dempster's rule of conditioning tells us how to modify a belief function Bel: $\mathcal{Q} \to [0, 1]$ when we learn that $A \in \mathcal{Q}$ is true. His more general rule of combination tells us how to modify Bel when the evidence underlying it is pooled with independent evidence underlying a second belief function Bel': $\mathcal{Q} \to [0, 1]$.

In section 7, we will see how the rule of combination provides a justification for the term "cognitively independent," which was introduced in the preceding chapter.

1. Dempster's Rule of Conditioning

The central feature of the theory of subjective probability is its rule of conditioning. The rule is open to criticism but it has a tremendous intuitive appeal and has always been accepted by students of subjective probability. In this section, I will describe the rule from an intuitive point of view and introduce the analogous rule for belief functions.

Suppose we are dealing with a set $\mathcal{S}$ which is the set of all possible values of some quantity $\underset{\sim}{s}$ whose true value is unknown, and suppose we have a probability function

$$P: \mathcal{P}(\mathcal{S}) \to [0, 1],$$

$P(S)$ being our degree of belief (or subjective probability) that the true value of $\underset{\sim}{s}$ is in S. Then we can describe our situation intuitively by saying that our probability is distributed over the set $\mathcal{S}$. Now suppose we learn, from new evidence, that the true value of $\underset{\sim}{s}$ is really in a proper subset $\mathcal{S}_o$ of $\mathcal{S}$. Then if $P(\mathcal{S}_o) < 1$ our probability function P will evidently require modification, for we will now wish to assert a degree of belief 1 in $\mathcal{S}_o$. How should P be modified?

The obvious thing to do is to "throw away" that portion of our probability that was distributed over $\mathcal{S}_o$; it was committed to something that is now seen as impossible, so it seems that the only thing that can be done is to discard it. This will leave us, of course, with a total amount of probability that has measure $P(\mathcal{S}_o)$, which may be less than one. Hence we will want to "renormalize" the measure of all our remaining probability, multiplying all the measures by $1/P(\mathcal{S}_o)$ so as to bring the measure of the total back up to one again.

This procedure will result in a new probability function P' over $\mathcal{S}$, one that now gives $P'(\mathcal{S}_o) = 1$. In order to describe this probability function explicitly, let us refer to Figure 1 and calculate the value of P' for each of the sets S_1, S_2 and S_3 shown there. First of all, all the probability that was committed to S_1 has been thrown away; hence we now have

$$P'(S_1) = 0. \tag{1}$$

As for S_2, none of the probability associated with it has been thrown away, but its measure has been renormalized, so we have

$$P'(S_2) = P(S_2) \;/\; P(\mathcal{S}_o). \tag{2}$$

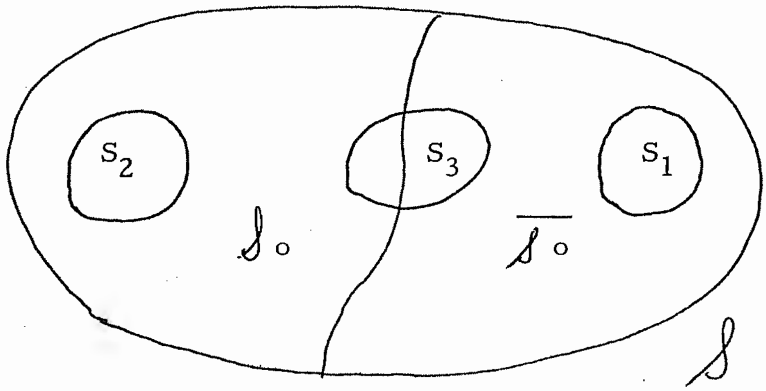

Figure 1.

Finally, consider S_3. Some of the probability that was distributed over S_3, namely the portion which was distributed over $S_3 \cap \overline{\mathcal{S}_o}$, has been eliminated. Hence the portion remaining is what was distributed over $S_3 \cap \mathcal{S}_o$, which did have measure $P(S_3 \cap \mathcal{S}_o)$ and now has measure

$$P'(S_3) = P(S_3 \cap \mathcal{S}_o)/P(\mathcal{S}_o). \tag{3}$$

An examination of (1), (2) and (3) shows that (1) and (2) are actually special cases of (3) , which is thus the general rule for conditioning P on $\mathcal{S}_o$.

The fact that P' is conditional on $\mathcal{S}_o$ is often indicated by denoting it by $P_{\mathcal{S}_o}$ or $P(\cdot \mid \mathcal{S}_o)$. In these notations, our rule becomes

$$P_{\mathcal{S}_o}(S) = P(S \cap \mathcal{S}_o)/P(\mathcal{S}_o)$$

or

$$P(S \mid \mathcal{S}_o) = P(S \cap \mathcal{S}_o) \;/\; P(\mathcal{S}_o) \tag{4}$$

for all $S \subset \mathcal{S}$. This is the classical rule for conditional probability; it is easily verified directly that $P(\cdot \mid \mathcal{S}_o)$ does indeed satisfy the axioms for probability functions, provided only that $P(\mathcal{S}_o) > 0$. Of course, if

$P(\mathcal{S}_o) = 0$ then our new knowledge that the true value of $\underset{\sim}{s}$ is in $\mathcal{S}_o$ is in direct contradiction with P, and the conditioning cannot be carried out.

An analogous rule applies, of course, to a probability function P on any Boolean algebra $\mathcal{Q}$. If $P(A) > 0$, then conditioning P on A yields a probability function $P(\cdot | A)$ on $\mathcal{Q}$ given by

$$P(B|A) = P(B \wedge A) \, / \, P(A) \tag{5}$$

for all $B \in \mathcal{Q}$.

The intuition behind this classical rule generalizes directly to the case of belief functions. For suppose we begin with a belief function

$$\mathrm{Bel}: \mathcal{P}(\mathcal{S}) \rightarrow [0, 1]$$

and then learn that the true value of $\underset{\sim}{s}$ is actually in $\mathcal{S}_o \subset \mathcal{S}$. What do we do? Well, we eliminate the probability that is committed to $\overline{\mathcal{S}_o}$ and renormalize the rest; the measure of the probability eliminated is Bel $(\overline{\mathcal{S}_o})$, so the measure of the remainder will be $1 - \mathrm{Bel}(\overline{\mathcal{S}_o})$ and the constant of renormalization will be $(1 - \mathrm{Bel}(\overline{\mathcal{S}_o}))^{-1}$. There is only one new idea that must be introduced: since our probability is allocated in a semi-mobile way over $\mathcal{S}$ rather than being distributed in a fixed way, we must recognize that the restriction to $\mathcal{S}_o$ may further restrict the mobility of some of our probability without eliminating it entirely. This means that some of our probability that was not committed to a set $S \subset \mathcal{S}$ may become committed to S by the restriction to $\mathcal{S}_o$. In fact, any probability that was committed to $S \cup \overline{\mathcal{S}_o}$ before will now be committed to S, unless it was committed to $\overline{\mathcal{S}_o}$ and hence must be eliminated. In general, then, the amount of probability committed to S after conditioning will be the measure of the probability previously committed to $S \cup \overline{\mathcal{S}_o}$

less the measure of the probability eliminated, or

$$\mathrm{Bel}(S \cup \overline{\mathcal{S}_o}) - \mathrm{Bel}(\overline{\mathcal{S}_o}).$$

But this must be renormalized, so we obtain

$$\mathrm{Bel}(S \mid \mathcal{S}_o) = \frac{\mathrm{Bel}(S \cup \overline{\mathcal{S}_o}) - \mathrm{Bel}(\overline{\mathcal{S}_o})}{1 - \mathrm{Bel}(\overline{\mathcal{S}_o})} \tag{6}$$

as our degree of belief in S conditional on $\mathcal{S}_o$.

As it turns out, this rule is stated more easily in terms of the upper probability functions. Indeed,

$$P^*(S \mid \mathcal{S}_o) = 1 - \mathrm{Bel}(\overline{S} \mid \mathcal{S}_o)$$

$$= 1 - \frac{\mathrm{Bel}(\overline{S} \cup \overline{\mathcal{S}_o}) - \mathrm{Bel}(\overline{\mathcal{S}_o})}{1 - \mathrm{Bel}(\overline{\mathcal{S}_o})}$$

$$= \frac{1 - \mathrm{Bel}(\overline{S} \cup \overline{\mathcal{S}_o})}{1 - \mathrm{Bel}(\overline{\mathcal{S}_o})}$$

$$= \frac{1 - \mathrm{Bel}(\overline{S \cap \mathcal{S}_o})}{1 - \mathrm{Bel}(\overline{\mathcal{S}_o})},$$

or

$$P^*(S \mid \mathcal{S}_o) = \frac{P^*(S \cap \mathcal{S}_o)}{P^*(\mathcal{S}_o)} \quad . \tag{7}$$

This is <u>Dempster's rule of conditioning.</u> It is easily verified that $P^*(\cdot \mid \mathcal{S}_o)$ does indeed satisfy the rules for upper probability functions, provided only that $P^*(\mathcal{S}_o) > 0$. If $P^*(\mathcal{S}_o) = 0$, then our new knowledge that the true value of $\underset{\sim}{s}$ is in $\mathcal{S}_o$ is in direct contradiction with P^*, and the conditioning cannot be carried out.

Dempster's rule of conditioning need not, of course, be restricted to upper probability functions on power sets; it can be applied to the conditioning of any upper probability function

$$P^*: \mathcal{A} \rightarrow [0, 1]$$

on any proposition $A \in \mathcal{A}$ such that $P^*(A) > 0$. The resulting conditional upper probability function $P^*(\cdot \mid A)$ is given by

$$P^*(B \mid A) = \frac{P^*(B \wedge A)}{P^*(A)} \qquad (8)$$

for all $B \in \mathcal{A}$. If P^* is actually a probability function, this rule reduces to (5), the classical rule of conditional probability.

There is a difficulty with the application of the classical rule, and the generalization (8) might seem to suffer from the same difficulty. The difficulty is that we sometimes feel that $P(A) = 0$ does not really mean that A is impossible. In the case of a "continuous" distribution of probability P over a set $\mathcal{S}$, for example, $P(\{s\}) = 0$ for every $s \in \mathcal{S}$; yet this is not taken to mean that it is impossible for the true value of $\underset{\sim}{s}$ to be s for every $s \in \mathcal{S}$. Hence in general it may be impossible to carry out the conditioning even in cases where we would like to do so. Interestingly enough, though, <u>condensable</u> belief functions are exempt from this difficulty. Indeed, when an upper probability function $P^*: \mathcal{P}(\mathcal{S}) \rightarrow [0, 1]$ is condensable we are entitled to interpret $P^*(S) = 0$ as meaning that P^* holds it to be impossible for the true value of $\underset{\sim}{s}$ to be in S. (See the end of section 1 of Chapter 5.) Hence our inability to condition a condensable upper probability on a set of upper probability zero need never be embarrassing, and the rule of conditioning appears to be most adapted to the condensable case.

It is easily verified that if $P^*: \mathcal{P}(\mathcal{S}) \to [0, 1]$ is condensable and $P^*(A) > 0$, then $P^*(\cdot \mid A)$ will also be condensable. And the commonality numbers are affected by conditioning in a very simple way. The commonality function Q for P^* is given, of course, by

$$Q(B) = - \sum_{T \subset B} (-1)^{\text{card } T} P^*(T)$$

for finite non-empty subsets B of $\mathcal{S}$. And the commonality function $Q(\cdot \mid A)$ for $P^*(\cdot \mid A)$ will be given by

$$Q(B \mid A) = - \sum_{T \subset B} (-1)^{\text{card } T} P^*(T \mid A)$$

$$= - \sum_{T \subset B} (-1)^{\text{card } T} \frac{P^*(T \cap A)}{P^*(A)}$$

$$= \frac{-1}{P^*(A)} \Big(\sum_{R \subset B \cap A} \sum_{S \subset B \cap \overline{A}} (-1)^{\text{card } R} (-1)^{\text{card } S} P^*((R \cup S) \cap A) \Big)$$

$$= \frac{-1}{P^*(A)} \Big(\sum_{R \subset B \cap A} (-1)^{\text{card } R} P^*(R) \Big) \Big(\sum_{S \subset B \cap \overline{A}} (-1)^{\text{card } S} \Big).$$

Now if $B \subset A$, then the last factor is equal to one; otherwise it is equal to zero. Hence

$$Q(B|A) = \begin{cases} 1 & \text{if } B = \phi \\ \dfrac{Q(B)}{P^*(A)} & \text{if } \phi \neq B \subset A, \\ 0 & \text{otherwise.} \end{cases}$$

So conditioning a condensable allocation can be carried out by renormalizing the relevant commonality numbers.

In the case of a belief function on a finite Boolean algebra $\mathcal{A}$, it is also possible to describe the conditioning process in terms of the basic probability numbers. Suppose indeed that Bel: $\mathcal{A} \rightarrow [0, 1]$ is given by the basic probability numbers $\{m_A\}_{A \in \mathcal{A}}$: Then upon conditioning on A, the basic probability mass that was associated with $A' \in \mathcal{A}$ will be constrained to $A' \wedge A$. Hence there will come to be associated with $B \in \mathcal{A}$ a total basic probability mass of measure

$$\Sigma \{m_{A'} | A' \wedge A = B\}.$$

In particular a basic probability mass of measure

$$\Sigma \{m_{A'} | A' \wedge A = \Lambda\} = \mathrm{Bel}(\overline{A})$$

will come to be associated with Λ. This latter probability mass must of course be eliminated, and we must renormalize by the factor $(P^*(A))^{-1}$, thus obtaining the new basic probability numbers $\{m'_B\}_{B \in \mathcal{A}}$ given by

$$m'_B = \frac{\Sigma \{ m_{A'} | A' \wedge A = B \}}{P^*(A)}$$

for all $B \neq \Lambda_{\mathcal{A}}$ and, of course, $m_\Lambda = 0$.

2. The Conditional Allocation

Dempster's rule of conditioning is most simply described intuitively in terms of mobile probability masses: in order to condition Bel: $\mathcal{Q} \to [0, 1]$ on $A \in \mathcal{Q}$, we add to the constraints on all our probability masses by constraining each one to A, and hence to $A \wedge A'$ for all $A' \in \mathcal{Q}$ to which it was previously constrained; we then eliminate all the probability that is constrained to Λ by this process. In order to represent this process mathematically, we must use the formal procedure that we learned in section 4 of Chapter 4 for "discarding" a probability mass from a probability algebra.

Theorem. Let $\rho: \mathcal{Q} \to \mathcal{M}$ be an allocation into the probability algebra $(\mathcal{M}, \mu)$. Suppose $A \in \mathcal{Q}$ and $\rho(\overline{A}) \neq \mathrm{V}$. Let I be the ideal in generated by $\rho(\overline{A})$, and let $(\mathcal{M}/I, \nu)$ be as in section 4 of Chapter 4. Let $f: \mathcal{M} \to \mathcal{M}/I$ be the canonical homomorphism. Then

$$\rho_A: \mathcal{Q} \to \mathcal{M}/I : A' \leadsto f(\rho\ (A' \vee \overline{A})\ )$$

is an allocation, and $\mathrm{Bel}_A = \nu \circ \rho_A$ is given by

$$\mathrm{Bel}_A(A') = \frac{\mathrm{Bel}(A' \vee \overline{A}) - \mathrm{Bel}(\overline{A})}{1 - \mathrm{Bel}(\overline{A})}$$

for all $A' \in \mathcal{Q}$.

Proof: It is easy to verify that ρ_A is an allocation:

(i) $\rho_A(\Lambda) = f(\rho(\overline{A})) = \Lambda$,

(ii) $\rho_A(\mathcal{T}) = f(\rho(\bar{A} \vee \mathcal{T})) = f(\mathcal{T}) = \mathcal{T}$,

(iii) $\rho_A(A_1 \wedge A_2) = f(\rho((A_1 \wedge A_2) \vee \overline{A})) = f(\rho((A_1 \vee \overline{A}) \wedge (A_2 \vee \overline{A})))$

$$= f(\rho(A_1 \vee \overline{A})) \wedge f(\rho(A_2 \vee \overline{A}))$$

$$= \rho_A(A_1) \wedge \rho_A(A_2).$$

And

$$\mathrm{Bel}_A(A') = \nu(f(\rho(A' \vee \overline{A}))) = \frac{1}{1 - \mu(\rho(\overline{A}))} \mu(\rho(A' \vee \overline{A}) - \rho(\overline{A}))$$

$$= \frac{\mathrm{Bel}(A' \vee \overline{A}) - \mathrm{Bel}(\overline{A})}{1 - \mathrm{Bel}(\overline{A})}$$

by the formula in section 4 of Chapter 4.

The allocation ρ_A is called, of course, the <u>conditional allocation</u> obtained from ρ by conditioning on A.

3. Two Examples of Conditioning

In this section I will illustrate Dempster's rule of conditioning with two simple examples.

A. The Senate Example

First let us reconsider the example from Chapter 1 that involved an allocation of probability over the set of twenty-two Senators. That set is pictured again in Figure 2. Recall that our allocation of probability involved eleven basic probability masses, one corresponding to each

Langdon	(D)	Wingate	(D)
Few	(D)	Gunn	(D)
Lee	(D)	Grayson	(D)
Izard	(D)	Butler	(D)
Johnson	(D)	Ellsworth	(F)
Maclay	(D)	Morris	(F)
Strong	(F)	Dalton	(F)
Paterson	(F)	Elmer	(F)
Bassett	(F)	Read	(F)
Carroll	(F)	Henry	(F)
King	(F)	Schuyler	(F)

Figure 2.

State, and that each of these is free to move back and forth between the two Senators from the State to which it corresponds. We concluded that the degree of belief and the upper probability for the proposition A = "A Democratic-Republican will be chosen" were given by Bel(A) = 4/11 and P*(A) = 6/11.

Now Senator Maclay of Pennsylvania was particularly well known as a staunch anti-Federalist. Let us suppose that we begin with the allocation of probability just described but that we then learn -- say from a friend galloping past who pauses only to mention the fact with a sigh of relief -- that Maclay was <u>not</u> chosen. After the receipt of this information, what degree of belief and upper probability ought we to accord to the proposition A?

Well, we must condition our allocation of probability to the set $\overline{\{\text{Maclay}\}}$, i.e., to the set of the twenty-one Senators other than Maclay. This conditioning will not eliminate any of our probability, and it will change the region of mobility of only one of the eleven basic probability masses. The basic probability mass corresponding to the State of Pennsylvania, instead of moving freely between Senators Maclay and Morris, will now be constrained to Senator Morris. Hence there will still be only four basic probability masses constrained to Democratic-Republican Senators, but six of the seven remaining ones will be constrained to Federalist Senators. So conditionally we will have a degre of belief of 4/11 for A but an upper probability of only 5/11.

B. Conditioning on the Diagonal

In section 1 of Chapter 6 we considered an example in which we began with a belief function

$$\text{Bel}_o: \mathcal{P}(\mathcal{J}_1) \to [0, 1],$$

which expressed our degrees of belief about the true value of an unknown quantity $\underset{\sim}{X}$, $\mathcal{J}_1$ being the set of possible values of $\underset{\sim}{X}$. We also considered a second unknown quantity $\underset{\sim}{Y}$, about the true value of which we had no opinions save that it was in $\mathcal{J}_2$; and we used Bel_o to obtain a belief function

$$\text{Bel}: \mathcal{P}(\mathcal{J}_1 \times \mathcal{J}_2) \to [0, 1],$$

which expressed our degrees of belief in joint propositions about the true values of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$. Bel was given by

$$\mathrm{Bel}(A) = \mathrm{Bel}_o(\{x \mid \{x\} \times \mathcal{S}_2 \subset A\}),$$

Bel(A) being our degree of belief that the pair consisting of the true value of $\underset{\sim}{X}$ and the true value of $\underset{\sim}{Y}$ was in A.

Now let us suppose that $\mathcal{S}_1$ and $\mathcal{S}_2$ are actually the same set: $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$; so that our belief function Bel is actually on $\mathcal{P}(\mathcal{S} \times \mathcal{S})$. Now suppose that we suddenly learn that the quantities $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are identical -- that they have the same value. Then how should be modify Bel?

Evidently, we should condition Bel on the "diagonal" -- on the set

$$D = \{(s,s) \mid s \in \mathcal{S}\}.$$

This does not result in the elimination of any probability, for

$$\mathrm{Bel}(\overline{D}) = \mathrm{Bel}_o\left(\{x \mid \{x\} \times \mathcal{S}_2 \subset \overline{\{(s,s) \quad s \in \mathcal{S}\}}\}\right)$$

$$= \mathrm{Bel}_o(\phi) = 0.$$

So the conditional belief function Bel_D is given simply by

$$\begin{aligned} \mathrm{Bel}_D(A) &= \mathrm{Bel}(A \vee \overline{D}) \\ &= \mathrm{Bel}_o(\{x \mid \{x\} \times \mathcal{S}_2 \subset A \vee \overline{D}\}) \\ &= \mathrm{Bel}_o(\{x \mid (x,x) \in A\}) \end{aligned}$$

We might be interested in particular in $\mathrm{Bel}_D \big| \mathcal{P}(\mathcal{S}_2)$, which would give our conditional degrees of belief that the true value of $\underset{\sim}{Y}$ is in various subsets of $\mathcal{S}_2 = \mathcal{S}$. Denoting this belief function by

$$\mathrm{Bel}' : \mathcal{P}(\mathcal{S}_2) \to [0,1],$$

we would have

$$\mathrm{Bel}'(A) = \mathrm{Bel}_D(\mathcal{S}_1 \times A) = \mathrm{Bel}_o(\{x \mid (x,x) \in \mathcal{S}_1 \times A\})$$

$$= \mathrm{Bel}_o(A).$$

Hence our conditioning has resulted in the same degrees of belief for $\underset{\sim}{Y}$ as we formerly had for $\underset{\sim}{X}$. Nothing could be more reasonable.

4. Dempster's Rule of Combination: Finite Case

Suppose we have two belief functions Bel_1 and Bel_2 on the same Boolean algebra $\mathcal{Q}$, and suppose the two are based on independent sources of evidence. Then it would be pleasant if we could combine them in some orthogonal way so as to produce a single resulting belief function on $\mathcal{Q}$; this would correspond to pooling the evidence from which the two belief functions arose. How might we carry out such a combination?

This question can be approached most easily in the case where $\mathcal{Q}$ is finite. In that case, it should be recalled, a belief function Bel on $\mathcal{Q}$ can be described by "basic probability numbers" $\{m_A\}_{A \in \mathcal{Q}}$. The intuitive understanding is that the basic probability number m_A represents the measure of a "basic probability mass" which is constrained to A but not to any proper subelement of A. Suppose we have two belief functions Bel_1 and Bel_2 on $\mathcal{Q}$, with basic probability numbers $\{n_A\}_{A \in \mathcal{Q}}$ and $\{p_A\}_{A \in \mathcal{Q}}$, respectively. In order to think about combining Bel_1 and Bel_2, let us think of Bel_1 as our own original belief function, while Bel_2 is the belief function of a second person whose opinions we wish to combine orthogonally with our own. How can be use Bel_2 to modify our original beliefs?

Well, let us consider each of the other person's basic probability masses separately. The basic probability mass which he associates with A is committed to A but to no proper subelement of A. As far as that probability mass is concerned, the natural thing seems to be to condition Bel_1 on A. In other words, we should restrict each of the basic probability masses for Bel_1 to A, thus obtaining a basic probability mass for each $B \in \mathcal{Q}$ of measure

$$\Sigma \{n_{A'} \mid A' \wedge A = B\}.$$

But this should apply only for Bel_2's basic probability mass for A, which has measure p_A. Doing to the same for each $A \in \mathcal{Q}$, we would obtain the total

$$\sum_{A \in \mathcal{Q}} p_A \; \Sigma \{ n_{A'} \mid A' \wedge A = B \}. \qquad (1)$$

as the measure of the new basic probability mass associated with β.

The difficulty with (1) is, of course, that it may be positive for $B = \Lambda$; there may be some probability that is constrained to Λ as a result of this rule. Hence we must discard that portion of our probability and renormalize the measure of the remainder. This results in a new belief function Bel with basic probability numbers $\{m_B\}_{B \in \mathcal{Q}}$, where

$$m_B = \frac{\sum_{A \in \mathcal{Q}} p_A \; \Sigma \{n_{A'} \mid A' \wedge A = B\}}{1 - \sum_{A \in \mathcal{Q}} p_A \; \Sigma \{n_{A'} \mid A' \wedge A = \Lambda\}} \qquad (2)$$

for $B \neq \Lambda$, and $m_{\Lambda} = 0$.

The numbers $\{m_B\}_{B \in \mathcal{Q}}$ defined by (2) are evidently non-negative, so in order to show that they determine a belief function it suffices to show that they add to one, and this is easily verified. The only difficulty that might arise is that we might have

$$\sum_{A \in} p_A \sum \{n_{A'} | A' \wedge A = \Lambda \} = 1; \tag{3}$$

in such a case the denominator in (2) would be zero and the combination could not be carried out. But since $\sum_{A \in} p_A = 1$, (3) would imply that

$$\sum \{n_{A'} | A' \wedge A = \Lambda \} = 1$$

for all A for which $p_A > 0$. Denoting the A for which $p_A > 0$ by $A_1, \ldots, A_k$, we find that

$$Bel_1(\overline{A_i}) = \sum \{ n_{A'} | A' \leq \overline{A_i} \} = \sum \{n_{A'} | A' \wedge A_i = \Lambda \} = 1$$

for each i, $i = 1, \ldots, k$. Setting $C = A_1 \vee \ldots \vee A_k$, this implies that

$$Bel_1(\overline{C}) = Bel_1(\overline{A_1} \wedge \ldots \wedge \overline{A}_k) = 1,$$

while

$$Bel_2(C) = \sum_{A \leq C} p_A = 1.$$

So the combination of Bel_1 and Bel_2 is impossible only when there exists $C \in \mathcal{Q}$ such that $P_1{}^*(C) = 0$ but $Bel_2(C) = 1$; i.e., when the two belief functions contradict each other.

5. Dempster's Rule of Combination: General Case

There are several approaches that we might take to adduce Dempster's rule of combination for the infinite case. One approach would be to develop the theory of integration for probability algebras and use it to adduce integrals analogous to the sums in formula (1) of the preceding chapter. An approach that we are better equipped to pursue is to draw an analogy with the "orthogonal combination" of Chapter 6, modifying that method by adding the element of renormalization. This is the approach of the following theorem.

Theorem. Suppose $Bel_1: \mathcal{A} \to [0,1]$ and $Bel_2: \mathcal{A} \to [0,1]$ are both belief functions, with standard representations $\rho_1^\Delta: \mathcal{A} \to \mathcal{M}_1$ and $\rho_2^\Delta: \mathcal{A} \to \mathcal{M}_2$. Let $((\mathcal{M}, \mu); i_1: \mathcal{M}_1 \to \mathcal{M}; i_2: \mathcal{M}_2 \to \mathcal{M})$ be an orthogonal sum of $(\mathcal{M}_1, \mu_1)$ and $(\mathcal{M}_2, \mu_2)$. Denote $\rho_1' = i_1 \circ \rho_1^\Delta$ and $\rho_2' = i_2 \circ \rho_2^\Delta$. And suppose that

$$M = \bigvee_{A \in \mathcal{A}} (\rho_1'(A) \wedge \rho_2'(\bar{A})) \neq V_{\mathcal{M}}.$$

Denote by I the principal ideal of $\mathcal{M}$ generated by M, and let $(\mathcal{M}/I, \nu)$ and $f: \mathcal{M} \to \mathcal{M}/I$ be as in section 4 of Chapter 4. Then

$$\rho': \mathcal{A} \to \mathcal{M}/I : A \leadsto f(\vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \in \mathcal{A}; A_1 \wedge A_2 \leq A \})$$

is a standard allocation of probability on $\mathcal{A}$. And the belief function $Bel = \nu \circ \rho'$ is given by

$$Bel(A) = \frac{k(A) - k}{1 - k},$$

where

$$k(A) = \sup \{ \Sigma \operatorname{Bel}_1(A_i) \operatorname{Bel}_2(B_i) - \Sigma \operatorname{Bel}_1(A_i \wedge A_j) \operatorname{Bel}_2(B_i \wedge B_j) +- \ldots +(-1)^{n+1} \operatorname{Bel}_1(A_1 \wedge \ldots \wedge A_n) \operatorname{Bel}_2(B_1 \wedge \ldots \wedge B_n) \mid n \geq 1;\ A_i, B_i \in \mathcal{Q};\ A_i \wedge B_i \leq A \} \quad (1)$$

and $k = k(\Lambda) = \mu(M)$

$$= \sup \{ \Sigma \operatorname{Bel}_1(A_i) \operatorname{Bel}_2(\overline{A_i}) - \Sigma \operatorname{Bel}_1(A_i \wedge A_j) \operatorname{Bel}_2(\overline{A_i} \wedge \overline{A_j}) +- \ldots + (-1)^{n+1} \operatorname{Bel}_1(A_1 \wedge \ldots \wedge A_n) \operatorname{Bel}_2(\overline{A_1} \wedge \ldots \wedge \overline{A_n}) \mid n \geq 1;\ A_1, \ldots, A_n \in \mathcal{Q} \}. \quad (2)$$

<u>Proof:</u> To show that ρ' is an allocation, notice that

(i) $\rho'(\Lambda) = f(\vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \in \mathcal{Q};\ A_1 \wedge A_2 \leq \Lambda \})$

$= f(M) = \Lambda,$

(ii) $\rho'(\mathcal{V}) = f(\vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \in \mathcal{Q};\ A_1 \wedge A_2 \leq \mathcal{V} \})$

$= f(\mathcal{V}) = \mathcal{V},$

and (iii) $\rho'(A) \wedge \rho'(B) = f(\vee \{ \rho_1'(R) \wedge \rho_2'(S) \mid R \wedge S \leq A \}) \wedge f(\vee \{ \rho_1'(T) \wedge \rho_2'(U) \mid T \wedge U \leq B \})$

$= f(\vee \{ \rho_1'(R) \wedge \rho_2'(S) \wedge \rho_1'(T) \wedge \rho_2'(U) \mid R \wedge S \leq A;\ T \wedge U \leq B \})$

$= f(\vee \{ \rho_1'(R \wedge T) \wedge \rho_2'(S \wedge U) \mid R \wedge S \leq A;\ T \wedge U \leq B \})$

$= f(\vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1 \wedge A_2 \leq A \wedge B \})$

$= \rho'(A \wedge B).$

Now by the formula in section 4 of Chapter 4,

$$\mathrm{Bel}(A) = \nu(f(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) | A_1 \wedge A_2 \leq A\}))$$

$$= \frac{1}{1-\mu(\mathcal{M})} \mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) | A_1 \wedge A_2 \leq A\} - M)$$

$$= \frac{1}{1-\mu(\mathcal{M})} \Big[\mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) | A_1 \wedge A_2 \leq A\})$$

$$- \mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) | A_1 \wedge A_2 \leq \Lambda\})\Big]$$

$$= \frac{1}{1-k}(k(A) - k),$$

where

$$k(A) = \mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) | A_1 \wedge A_2 \leq A\})$$

$$= \mu(\vee\{[\rho_1'(A_1) \wedge \rho_2'(B_1)] \vee [\rho_1'(A_2) \wedge \rho_2'(B_2)] \vee \ldots \vee$$

$$[\rho_1'(A_n) \wedge \rho_2'(B_n)] | A_i \wedge B_i \leq A \text{ for each } i\})$$

$$= \sup\{\mu([\rho_1'(A_1) \wedge \rho_2'(B_1)] \vee \ldots \vee [\rho_1'(A_n) \wedge \rho_2'(B_n)]) |$$

$$A_i \wedge B_i \leq A \text{ for each } i\}$$

$$= \sup\{\Sigma\, \mathrm{Bel}_1(A_i)\, \mathrm{Bel}_2(B_i) - \Sigma\, \mathrm{Bel}_1(A_i \wedge A_j)\, \mathrm{Bel}_2$$

$$(B_i \wedge B_j) +- \ldots + (-1)^{n+1}\, \mathrm{Bel}_1(A_1 \wedge \ldots \wedge A_n)$$

$$\mathrm{Bel}_2(B_1 \wedge \ldots \wedge B_n) \,|\, n \geqq 1;\ A_i, B_i \in \mathcal{Q};\ A_i \wedge B_i \leq A\},$$

and $k = k(\Lambda)$

$$= \mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) | A_1 \wedge A_2 \leq \Lambda\})$$

$$= \mu(\bigvee_{A \in \mathcal{Q}} \{\rho_1'(A) \wedge \rho_2'(\bar{A})\})$$

$$= \sup \{\Sigma \mathrm{Bel}_1(A_i)\ \mathrm{Bel}_2(\overline{A}_i) - \Sigma \mathrm{Bel}_1(A_i \wedge A_j)\ \mathrm{Bel}_2(\overline{A}_i \wedge \overline{A}_j) + - \ldots + (-1)^{n+1}\ \mathrm{Bel}_1(A_1 \wedge \ldots \wedge A_n) \cdot \mathrm{Bel}_2(\overline{A}_1 \wedge \ldots \wedge \overline{A}_n) \mid n \geq 1;\ A_1, \ldots, A_n \in \mathcal{A}.\}.$$

Definition. Suppose Bel_1 and Bel_2 are two belief functions on a Boolean algebra $\mathcal{A}$. If k, as given by (2) above, obeys $k < 1$, then the belief function Bel defined in the above theorem is called the orthogonal sum of Bel_1 and Bel_2 and is denoted $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$. If $k = 1$, then the orthogonal sum of Bel_1 and Bel_2 is said not to exist.

Notice that the formulae giving the orthogonal sum do not depend on the particular representations ρ_1', ρ_2' and ρ'.

The preceding is a definition of "orthogonal sum" in the case of two belief functions on the same Boolean algebra. But in the preceding chapter we defined the notion of an orthogonal sum of two belief functions on different independent subalgebras of a Boolean algebra. The following theorem shows in what sense the present definition is a generalization of the previous definition.

Theorem. Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of a Boolean algebra $\mathcal{A}$. And suppose $\mathrm{Bel}_1 : \mathcal{A}_1 \to [0, 1]$ and $\mathrm{Bel}_2 : \mathcal{A}_2 \to [0, 1]$ are belief functions. Denote by Bel_1' and Bel_2' the natural extensions of Bel_1 and Bel_2, respectively to $\mathcal{A}$. Let $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ be the orthogonal sum of Bel_1 and Bel_2 on $\mathcal{A}$, as defined in the preceding chapter. And let $\mathrm{Bel}_1' \oplus \mathrm{Bel}_2'$ be the orthogonal sum of Bel_1' and Bel_2', as defined above. Then $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2 = \mathrm{Bel}_1' \oplus \mathrm{Bel}_2'$.

<u>Proof</u>: Let $\rho_1^{\circ}: \mathcal{A}_1 \to \mathcal{M}_1$ and $\rho_2^{\circ}: \mathcal{A}_2 \to \mathcal{M}_2$ be as in the first theorem of section 3 of Chapter 3. Let $((\mathcal{M}, \mu); i_1: \mathcal{M}_1 \to \mathcal{M}; i_2: \mathcal{M}_2 \to \mathcal{M})$, ρ_1 and ρ_2 and ρ be as in that theorem as well. Then $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2 = \mu \circ \rho$, where

$$\rho = \vee \{ (\rho_1(A_1) \wedge \rho_2(A_2) \mid A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \le A \}.$$

But $\mathrm{Bel}_1': \mathcal{A} \to [0,1]$ and $\mathrm{Bel}_2': \mathcal{A} \to [0,1]$ are given by the allocations

$$\rho_1^{\Delta}: \mathcal{A} \to \mathcal{M}_1 : A \rightsquigarrow \vee \{ \rho_1^{\circ}(A_1) \mid A_1 \in \mathcal{A}_1;\ A_1 \le A \}$$

and

$$\rho_2^{\Delta}: \mathcal{A} \to \mathcal{M}_2 : A \rightsquigarrow \vee \{ \rho_2^{\circ}(A_2) \mid A_2 \in \mathcal{A}_2;\ A_2 \le A \}.$$

So $\mathrm{Bel}_1' \oplus \mathrm{Bel}_2'$ is given by

$$\rho': \mathcal{A} \to \mathcal{M}/I : A \rightsquigarrow f(\vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \le A \}),$$

where $\rho_1' = i_1 \circ \rho_1^{\Delta}$, $\rho_2' = i_2 \circ \rho_2^{\Delta}$ and

$$\begin{aligned}
I &= \bigvee_{A \in \mathcal{A}} (\rho_1'(A) \wedge \rho_2'(\bar{A})) \\
&= \bigvee_{A \in \mathcal{A}} (i_1(\vee \{ \rho_1^{\Delta}(A_1) \mid A_1 \in \mathcal{A}_1;\ A_1 \le A \}) \wedge i_2(\vee \{ \rho_2^{\Delta}(A_2) \mid \\
&\qquad A_2 \in \mathcal{A}_2;\ A_2 \le A \})) \\
&= \bigvee_{A \in \mathcal{A}} (\vee \{ i_1(\rho_1^{\Delta}(A_1)) \wedge i_2(\rho_2^{\Delta}(A_2)) \mid A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2; \\
&\qquad A_1 \le A;\ A_2 \le \bar{A} \}) \\
&= \perp,
\end{aligned}$$

since $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent. Hence $(\mathcal{M}/I, \nu) = (\mathcal{M}, \mu)$ and f is the identity mapping. And ρ' is given by

$$\begin{aligned}
\rho'(A) &= \vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq A\} \\
&= \vee\{i_1(\vee\{\rho_1^{\Delta}(A_1') \mid A_1 \in \mathcal{A}_1;\ A_1' \leq A_1\}) \wedge i_2(\vee\{\rho_2^{\Delta}(A_2') \mid \\
&\quad A_2' \in \mathcal{A}_2;\ A_2' \leq A_2\}) \mid A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq A\} \\
&= \vee\{(\vee\{\rho_1(A_1') \mid A_1 \in \mathcal{A}_1;\ A_1' \leq A_1\}) \wedge (\vee\{\rho_2(A_2') \mid A_2 \in \mathcal{A}_2; \\
&\quad A_2' \leq A_2\}) \mid A_1,\ A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq A\} \\
&= \vee\{\vee\{\rho_1(A_1') \wedge \rho_2(A_2') \mid A_1 \in \mathcal{A}_1;\ A_1' \leq A_1;\ A_2' \leq A_2\} \mid \\
&\quad A_1,\ A_2 \in \mathcal{A},\ A_1 \wedge A_2 \leq A\} \\
&= \vee\{\rho_1(A_1) \wedge \rho_2(A_2) \mid A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \leq A\} \\
&= \rho(A).
\end{aligned}$$

So $\rho' = \rho$, and hence $Bel_1' \oplus Bel_2' = Bel_1 \oplus Bel_2$. ▨

So our present notion of combination is quite general. Of course, one can combine more than two belief functions at a time; the more general definition should be obvious. The operation of combination is commutative whenever it can be carried out, and it has a unit -- the vacuous belief function -- which when combined with any belief function always yields that belief function again. The operation of conditioning is also a special case of combination, as the following theorem shows:

Theorem. Suppose $\mathrm{Bel}: \mathcal{Q} \to [0,1]$ is a belief function, $A \in \mathcal{Q}$ and $P^*(A) > 0$. Let $\mathrm{Bel}_2: \mathcal{Q} \to [0,1]$ be the belief function defined by

$$\mathrm{Bel}_2(A') = \begin{cases} 1 & \text{if } A \le A' \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mathrm{Bel}_A = \mathrm{Bel} \oplus \mathrm{Bel}_2$.

Proof: If $A' \in \mathcal{Q}$, then

$$\mathrm{Bel}_A(A') = \frac{\mathrm{Bel}(A' \vee \bar{A}) - \mathrm{Bel}(\bar{A})}{1 - \mathrm{Bel}(\bar{A})}$$

Now if we let ρ_1', ρ_2' and $(\mathcal{M}, \mu)$ be as in the first theorem of this section, we have $\mathrm{Bel} = \mu \circ \rho_1'$, and ρ_2' is given by

$$\rho_2'(A') = \begin{cases} \Upsilon & \text{if } A \le A' \\ \Lambda & \text{otherwise.} \end{cases}$$

Hence

$$\mathrm{Bel} \oplus \mathrm{Bel}_1(A') = \frac{k(A') - k}{1 - k} \quad ,$$

where

$$k = \mu\left(\bigvee_{A' \in \mathcal{Q}} (\rho_1'(A') \wedge \rho_2'(\bar{A}'))\right)$$

$$= \mu\left(\vee \{\rho_1'(A') \mid A \le \bar{A}'\}\right)$$

$$= \mu(\rho_1'(\bar{A})) = \mathrm{Bel}(\bar{A}),$$

and

$$k(A') = \mu\left(\vee \{\rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1 \wedge A_2 \le A'\}\right)$$

$$= \mu\left(\vee \{\rho_1'(A_1) \mid A_1, A_2 \in \mathcal{Q};\ A_1 \wedge A_2 \le A';\ A \le A_2\}\right)$$

$$= \mu(\rho_1'(A' \vee \bar{A})) = \mathrm{Bel}(A' \vee \bar{A}) .$$

▨

6. The Condensable Case

In the previous section we saw that Dempster's rule of combination could be adduced for belief functions in general. But in fact, this rule, like the rule of conditioning, is most adapted to the condensable case. For in that case the rule can be stated quite simply in terms of the commonality numbers, and it will fail only when the belief functions contradict each other.

Theorem. Suppose Bel_1 and Bel_2 are condensable belief functions on $\mathcal{P}(\mathcal{S})$. Then $Bel_1 \oplus Bel_2$ fails to exist if and only if there exists $S \subset \mathcal{S}$ such that $Bel_1(S) = Bel_2(\bar{S}) = 1$. And in the case where $Bel_1 \oplus Bel_2$ does exist, it is condensable, and its commonality function Q is given by

$$Q(S) = \frac{1}{1 - k} \; Q_1(S)\, Q_2(S) \tag{1}$$

for all finite non-empty subsets $S \subset \mathcal{S}$, where Q_1 and Q_2 are the commonality functions for Bel_1 and Bel_2, respectively, and k is the constant given in the first theorem of section 5.

Proof: This theorem is most easily established by comparing the construction in section 5 with the construction in section 3 of Chapter 6.

Think of Bel_1 and Bel_2 as belief functions on $\mathcal{P}(\mathcal{S}_1)$ and $\mathcal{P}(\mathcal{S}_2)$, respectively, where $\mathcal{S}_1$ and $\mathcal{S}_2$ are distinct copies of $\mathcal{S}$. Let $\rho_1^{\Delta}: \mathcal{P}(\mathcal{S}) \to \mathcal{M}_1$, $\rho_2^{\Delta}: \mathcal{P}(\mathcal{S}) \to \mathcal{M}_2$, $\mathcal{M}$, i_1, i_2, ρ_1', ρ_2', M and ρ' be as in the theorem of section 5. And let ρ_1^{o} and ρ_2^{o} be identical with ρ_1^{Δ} and ρ_2^{Δ} respectively, except that they are thought of as being on the copies $\mathcal{P}(\mathcal{S}_1)$ and $\mathcal{P}(\mathcal{S}_2)$, respectively. Let ρ_1, ρ_2 and ρ be the allocations based on ρ_1° and ρ_2° according to the formulae in section 3 of Chapter 6.

Let

$$D = \{ (s, s) \mid s \in \mathcal{S} \} \subset \mathcal{S}_1 \times \mathcal{S}_2 .$$

Then

$$\begin{aligned} \rho(\bar{D}) &= \vee \{ \rho_1(A_1) \wedge \rho_2(A_2) \mid A_1 \subset \mathcal{S}_1;\ A_2 \subset \mathcal{S}_2;\ A_1 \times A_2 \subset \bar{D} \} . \\ &= \vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \subset \mathcal{S};\ A_1 \cap A_2 = \phi \} \\ &= \vee \{ \rho_1'(A) \wedge \rho_2'(\bar{A}) \mid A \subset \mathcal{S} \} \\ &= M, \end{aligned}$$

and in general, for all $S \subset \mathcal{S}$, if $S' = \{ (s, s) \mid s \in S \} \subset \mathcal{S}_1 \times \mathcal{S}_2$, then

$$\begin{aligned} \rho(S' \cup \bar{D}) &= \vee \{ \rho_1(A_1) \wedge \rho_2(A_2) \mid A_1 \subset \mathcal{S}_1;\ A_2 \subset \mathcal{S}_2;\ A_1 \times A_2 \subset S' \cup \bar{D} \} \\ &= \vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \subset \mathcal{S};\ A_1 \cap A_2 \subset S \}. \end{aligned}$$

By comparing the theorem in section 2 with the first theorem in section 5, we see that ρ' is obtained from ρ by conditioning on D and then identifying D with $\mathcal{S}$ by the mapping $(s, s) \rightsquigarrow s$.

Hence our formula (1) becomes transparent; the multiplication follows from the similar rule in section 7 of Chapter 6, while the constant $1/(1 - k)$ results from the conditioning. ▨

7. Cognitive Independence

In the preceding chapter I suggested that two subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$ of a Boolean algebra $\mathcal{A}$ deserved to be called cognitively independent with respect to a belief function Bel on $\mathcal{A}$ if

$$P^*(A_1 \wedge A_2) = P^*(A_1) \cdot P^*(A_2) \tag{1}$$

for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. We are now in a position to examine the basis of that suggestion.

What ought we to mean when we say that two subalgebras are cognitively independent with respect to our opinions? Intuitively, we ought to mean that the assimilation of new evidence or opinion about the propositions in one of them would not change our degrees of belief in the propositions in the other. But Dempster's rules of conditioning and combination provide us with a mathematical representation of how new evidence or opinion can be assimilated, and hence we can make this intuitive understanding mathematically precise.

Indeed, if our new evidence about $\mathcal{A}_1$ comes down to the knowledge that $A_1 \in \mathcal{A}_1$ is true, then we would modify Bel by conditioning it on A_1. And, more generally, if our new evidence induced a belief function Bel_1 on $\mathcal{A}_1$, then we would modify Bel by replacing it with $Bel_1' \oplus Bel$, where Bel_1' is the natural extension of Bel_1 to $\mathcal{A}$. And as the following theorems show, these sorts of modifications in Bel will always fail to modify the degrees of belief in elements of $\mathcal{A}_2$ if and only if (1) holds.

Theorem. Suppose Bel: $\mathcal{A} \to [0,1]$ is a belief function and $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of $\mathcal{A}$. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are cognitively independent with respect to Bel if and only if

$\mathrm{Bel}_{A_1} | \mathcal{A}_2 = \mathrm{Bel} | \mathcal{A}_2$ whenever $A_1 \in \mathcal{A}_1$ and $P^*(A_1) > 0$.

Proof: $\mathrm{Bel}_{A_1} | \mathcal{A}_2 = \mathrm{Bel} | \mathcal{A}_2$ whenever $A_1 \in \mathcal{A}_1$ and $P^*(A_1) > 0$ if and only if

$$P^*(A_2) = \frac{P^*(A_2 \wedge A_1)}{P^*(A_1)}$$

for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ such that $P^*(A_1) > 0$. But this equation holds for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ such that $P^*(A_1) > 0$ if and only if

$$P^*(A_1 \wedge A_2) = P^*(A_1) \cdot P^*(A_2)$$

for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. ▨

Theorem. Suppose $\mathrm{Bel}: \mathcal{A} \to [0,1]$ is a belief function and $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of $\mathcal{A}$. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are cognitively independent with respect to Bel if and only if $\mathrm{Bel}_1' \oplus \mathrm{Bel} | \mathcal{A}_2 = \mathrm{Bel} | \mathcal{A}_2$ whenever Bel_1' is the natural extension to of a belief function Bel_1 on $\mathcal{A}_1$ and $\mathrm{Bel}_1' \oplus$ Bel exists.

Proof. In view of the preceding theorem, it suffices to show that (1) and the existence of $\mathrm{Bel}_1' \oplus \mathrm{Bel}$ implies that

$$(\mathrm{Bel}_1' \oplus \mathrm{Bel})(A) = \mathrm{Bel}(A)$$

for all $A \in \mathcal{A}_2$. But by the formulae of section 5, we find that

$$(\mathrm{Bel}_1' \oplus \mathrm{Bel})(A) = \Big(\mu(\vee\{\rho_1(A_1) \wedge \rho_2(A_2) \,|\, A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq A\}) - \mu(\vee\{\rho_1(A_1) \wedge \rho_2(A_2) \,|\, A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq \Lambda\})\Big) \Big/ \Big(1 - \mu(\vee\{\rho_1(A_1) \wedge \rho_2(A_2) \,|\, A_1, A_2 \in \mathcal{A};\ A \wedge A_2 \leq \Lambda\})\Big)$$

where $\rho_1 : \mathcal{Q} \to \mathcal{M}$ and $\rho_2 : \mathcal{Q} \to \mathcal{M}$ are allocations which represent Bel_1' and Bel, respectively, and which map $\mathcal{Q}$ into orthogonal subalgebras of $\mathcal{M}$. Now Bel_1' is supported by $\mathcal{Q}_1$; hence

$$\rho_1(A_1) = \vee \{ \rho_1(A') \mid A' \in \mathcal{Q}_1;\ A' \le A_1 \}$$

for all $A_1 \in \mathcal{Q}$, and it follows that

$$(Bel_1' \oplus Bel)(A) = \Big(\mu(\vee\{\rho_1(A_1) \wedge \rho_2(A \vee \bar{A}_1) \mid A_1 \in \mathcal{Q}_1\}) - \mu(\vee\{\rho_1(A_1) \wedge \rho_2(\bar{A}_1) \mid A_1 \in \mathcal{Q}_1\})\Big) \Big/ \Big(1 - \mu(\vee\{\rho_1(A_1) \wedge \rho_2(\bar{A}_1) \mid A_1 \in \mathcal{Q}_1\})\Big). \tag{2}$$

But

$$\mu(\vee\{\rho_1(A_1) \wedge \rho_2(A \vee \bar{A}_1) \mid A_1 \in \mathcal{Q}_1\})$$

$$= \sup_{A_1, \ldots, A_n \in \mathcal{Q}_1} \mu((\rho_1(A_1) \wedge \rho_2(A \vee \bar{A}_1)) \vee \ldots \vee (\rho_1(A_n) \wedge \rho_2(A \vee \bar{A}_n)))$$

$$= \sup_{A_1, \ldots, A_n \in \mathcal{Q}_1} \Big[\sum_i \mu(\rho_1(A_i) \wedge \rho_2(A \vee \bar{A}_i)) - \sum_{i<j} \mu(\rho_1(A_i \wedge A_j) \wedge \rho_2(A \vee (\bar{A}_i \wedge \bar{A}_j))) + - \ldots\Big]$$

$$= \sup_{A_1, \ldots, A_n \in \mathcal{Q}_1} \Big[\sum_i Bel_1(A_i)\, Bel(A \vee \bar{A}_i) - \sum_{i<j} Bel_1(A_i \wedge A_j)\, Bel(A \vee (\bar{A}_i \wedge \bar{A}_j)) + - \ldots\Big].$$

Now since $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are cognitively independent with respect to Bel, we have

$$P^*(A \wedge B) = P^*(A) \cdot P^*(B),$$

or

$$\mathrm{Bel}(A \vee B) = \mathrm{Bel}(A) + \mathrm{Bel}(B) - \mathrm{Bel}(A)\,\mathrm{Bel}(B)$$

whenever $A \in \mathcal{A}_2$ and $B \in \mathcal{A}_1$. We are indeed assuming that $A \in \mathcal{A}_2$, so our preceding formula becomes

$$\mu(\vee\{\rho_1(A_1) \wedge \rho_2(A \vee \overline{A}_1) \mid A_1 \in \mathcal{A}_1\})$$

$$= \sup_{A_1,\dots,A_n \in \mathcal{A}_1} \Big[\sum_i \mathrm{Bel}_1(A_i)\big((\mathrm{Bel}(A) + \mathrm{Bel}(\overline{A}_i) - \mathrm{Bel}(A)\,\mathrm{Bel}(\overline{A}_i)\big)$$

$$- \sum_{i<j} \mathrm{Bel}_1(A_i \wedge A_j)\big((\mathrm{Bel}(A) + \mathrm{Bel}(\overline{A}_i \wedge \overline{A}_j) -$$

$$- \mathrm{Bel}(A)\,\mathrm{Bel}(\overline{A}_i \wedge A_j)\big) + - \dots \Big]$$

$$= \sup_{A_1,\dots,A_n \in \mathcal{A}_1} \Big[\mathrm{Bel}(A)\,\big(\Sigma\,\mathrm{Bel}_1(A_i) - \Sigma\,\mathrm{Bel}_1(A_i \wedge A_j) + - \dots\big)$$

$$+ (1 - \mathrm{Bel}(A))\,(\Sigma\,\mathrm{Bel}_1(A_i)\,\mathrm{Bel}(\overline{A}_i) -$$

$$- \Sigma\,\mathrm{Bel}_1(A_i \wedge A_j)$$

$$\mathrm{Bel}(\overline{A}_i \wedge \overline{A}_j) + - \dots)\Big]$$

$$= \sup_{A_1,\dots,A_n \in \mathcal{A}_1} \Big[\mathrm{Bel}(A)\,(\mu(\rho_1(A_1) \vee \dots \vee \rho_1(A_n)))$$

$$+ (1 - \mathrm{Bel}(A))\,(\mu((\rho_1(A_1) \wedge \rho_2(\overline{A}_1)) \vee \dots \vee$$

$$(\rho_1(A_n) \wedge \rho_2(\overline{A}_n))))\Big]$$

$$= \mathrm{Bel}(A) \sup_{A_1,\dots,A_n \in \mathcal{A}_1} \mu(\rho_1(A_1) \vee \dots \vee \rho_1(A_n))$$

$$+ (1 - \mathrm{Bel}(A)) \sup_{A_1,\dots,A_n \in \mathcal{A}_1} \mu((\rho_1(A_1) \wedge \rho_2(\overline{A}_1)) \vee \dots \vee$$

$$(\rho_1(A_n) \wedge \rho_2(\overline{A}_n)))$$

$$= \mathrm{Bel}(A) + (1 - \mathrm{Bel}(A))\,\mu(\vee\{\rho_1(A_1) \wedge \rho_2(\overline{A}_1) \mid A_1 \in \mathcal{A}_1\}).$$

So, setting

$$k = \mu (\vee \{ \rho_1(A_1) \wedge \rho_2(\bar{A}_1) \mid A_1 \in \mathcal{Q}_1 \}),$$

(2) becomes

$$(\mathrm{Bel}_1{}' \oplus \mathrm{Bel})(A) = \frac{\mathrm{Bel}(A) + (1 - \mathrm{Bel}(A))\,k - k}{1 - k}$$

$$= \mathrm{Bel}(A).$$

8. Conclusion

It is evident that Dempster's rule of combination will play a central role in any application of the theory of belief functions, for we always encounter the need to combine evidence. In view of this importance, the rule deserves a much closer scrutiny -- we need to examine a good many examples of its application so as to understand its behavior clearly. I cannot undertake such an examination here, but I have made some efforts to examine its behavior in the paper entitled "A Theory of Statistical Support."

I have not developed the formulae for combining more than one belief function at a time, but it should be evident that such combination is possible. Furthermore, it can be carried out stepwise, and the order will not matter: the operation of combination is commutative. This is particularly obvious in the condensable case, for aside from an appropriate renormalization, the combination of condensable belief functions is affected merely by multiplying the commonality functions.

It should be noted that this operation of combination is <u>not</u> idempotent. In other words, Bel $\oplus$ Bel is not, in general, equal to Bel. This fact is best explicated if we think in terms of the evidence underlying Bel. Since the operation of combination corresponds to the pooling of evidence, Bel $\oplus$ Bel will be appropriate for the situation where all the evidence is twice as strong as that underlying Bel.

It is not so easy, of course, to go back and forth from the commonality functions, which are easy to manipulate, to the belief functions and upper probability functions, which are of greater immediate interest; the formulae for doing so that were adduced in Chapter 5 are hardly of practical use. Hence any application of this theory will involve the rather difficult task of developing effective computational methods for combination. This difficulty is central in the theory of Dempsterian inference, for which the present essay is meant as a foundation.

BIBLIOGRAPHY

Bartle, Robert G. (1966). The Elements of Integration. Wiley, New York.

Bernoulli, Daniel. (1777). "The most probable choice between several discrepant observations and the formation therefrom of the most likely induction." Acta Acad. Petrop. pp. 3-33. Reprinted as pp. 157-167 of Pearson and Kendall's Studies in the History of Statistics and Probability, Griffin, 1970.

Bernoulli, James. (1713). Artis Conjectandi. Basel. Especially pp. 217-223. See also pp. 22-34 of the translation by Bing Sung, issued as Technical Report No. 2 of the Department of Statistics, Harvard University, February 12, 1966.

Birkhoff, Garrett. (1967). Lattice Theory, Third Edition. American Mathematical Society.

Choquet, Gustave. (1953-4). "Theory of Capacities." Annales de l'Institut Fourier, Université de Grenoble, V. Pp. 131-296.

Dempster, A.P. (1966). "New methods for reasoning towards posterior distributions based on sample data." Ann. Math. Statist. 37, pp. 355-374.

Dempster, A.P. (1967). "Upper and lower probabilities induced by a multi-valued mapping." Ann. Math. Statist. 38, pp. 325-339.

Dempster, A.P. (1967). "Upper and lower probability inferences based on a sample from a finite univariate population." Biometrika 54, pp. 515-528.

Dempster, A.P. (1968). "Upper and lower probabilities generated by a random closed interval." Ann. Math. Statist. 39, pp. 957-966.

Dempster, A.P. (1968). "A generalization of Bayesian inference (with discussion)." J. Roy. Statist. Soc. Ser. B. 30, pp. 205-247.

Dempster, A.P. (1968). The Theory of Statistical Inference: A Critical Analysis. Chapter 1. Research Report S-2, Department of Statistics, Harvard University, August 26, 1968.

Dempster, A.P. (1968). The Theory of Statistical Inference: A Critical Analysis. Chapter 2. Research Report S-3, Department of Statistics, Harvard University, September 27, 1968.

Dempster, A.P. (1969). "Upper and lower probability inferences for families of hypotheses with monotone density ratios." Ann. Math. Statist. 40, pp. 953-969.

DePauw, Linda Grant, Ed. (1972). Documentary History of the First Federal Congress of the United States of America. Volume I: Senate Legislative Journal. The John Hopkins University Press. (Library of Congress Catalogue Number 73-155164.)

Diderot, Dénis. (1785). "Probabilité." Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers. Vol. 27, pp. 443-454.

Halmos, Paul R. (1963). Lectures on Boolean Algebras. Van Nostrand, New York.

Halmos, Paul R. (1950). Measure Theory. Van Nostrand, New York.

Halmos, Paul R. (1960). Naive Set Theory. Van Nostrand, New York.

Kappos, Demetrios A. (1969). Probability Algebras and Stochastic Spaces. Academic Press, New York.

Kolmogorov, A.N. and S.V. Fomin. (1961). Elements of the Theory of Functions and Functional Analysis. Volume 2. Graylock, Baltimore.

Lambert, Johann Heinrich. (1764). Neues Organon, Zweiter Band. Pp. 318-421. Reprinted in 1965 as Vol. II of Lambert's Philosophische Schriften by Georg Olms Verlagsbuchhandlung, Hildesheim.

Prevost and Lhuilier. (1797). "Mémoire sur l'application du Calcul des probabilités à la valeur du témoignage." Mémoires de l'Académie Royale de Berlin. Pp 120-152.

Shafer, Glenn. (1972). The Problem of Statistical Support. Department of Statistics, Princeton University. (This seminar paper is Part I of the larger work of which the present essay is Part II.)

Shafer, Glenn. (1973). A Theory of Statistical Support. An essay presented at a symposium on the foundations of probability and statistics at the University of Western Ontario, May 13, 1973.

Sikorski, Roman. (1969). Boolean Algebras. Springer-Verlag, New York.

Todhunter, Issac. (1865). A History of the Mathematical Theory of Probability. Reprinted by the Chelsea Publishing Co., New York, 1949.