THE CONCEPT OF DISTINCT EVIDENCE

Ph. SMETS
IRIDIA
Université Libre de Bruxelles
Brussels-Belgium

Abstract: In Dempster-Shafer theory, belief functions induced by distinct pieces of evidence can be combined by Dempster's rule of combination. The concept of distinctness has not been formally defined. We present a tentative definition of the concept of distinctness and compare this definition with the definition of stochastic independence described in probability theory.

1. Introduction.

Shafer (1978) introduced the idea that belief functions induced by "distinct" pieces of evidence should be combined by Dempster's rule of combination. But no definition of distinctness was provided, what lead to many misuses of Dempster's rule of combination (see Pearl (1990) for a list of errors and Smets (1991a) for solutions to these errors).

We present a tentative definition of the concept of distinctness and argue for its naturalness by comparing it to the concept of independence in probability theory. Our definition is coined within the transferable belief model, our interpretation of Dempster-Shafer theory (Smets 1991b, Smets and Kennes1990).

Our presentation is done under the open-world assumption (Smets 1988). It means that we do not require \( \text{bel}(\Omega)=1 \) or equivalently \( m(\emptyset)=0 \). Further we never normalize belief functions after conditioning or combination. The meaning of \( m(\emptyset)>0 \) is presented in Smets (1992).

2. Expansion is Specialization.

The TBM postulates the existence of basic belief masses (bbm) allocated to the subsets of a frame of discernment \( \Omega \). For \( A \subseteq \Omega \), the bbm \( m(A) \) quantifies the part of Your belief that supports \( A \) without supporting any strict subset of \( A \), and that could be transferred to subsets of \( A \) if further information justifies such a transfer. We call bba the function whose values on \( \Omega \) are the basic belief masses.

---

1 This research has been supported by the ESPRIT II Basic Research Project 3085 (DRUMS) and DGXII MUNVAR project, which are funded by the European Community.
2 'You' is the agent that entertains the beliefs considered in this presentation.
When a new piece of evidence becomes known to You, Your beliefs must be updated. There are three forms of updating: expansion, contraction and revision (Gardenfors, 1988). The **expansion** is the change of belief that results from adding a belief without retracting any old beliefs. The **contraction** is the change of belief that results from giving up a belief. It is the inverse of an expansion. Finally the **revision** is the change of belief that results from adding a belief that is not compatible with the previously held beliefs, in which case a contraction is also needed.

In the case of an expansion, each bbb \( m(A) \) for \( A \subseteq \Omega \) is distributed by the updating process among the subset of \( A \) (including the empty set \( \emptyset \)). This transfer of belief due to a new information can be described by a **specialization matrix** \( S \) defined on \( 2^{\Omega} \times 2^{\Omega} \) (Kruse and Schwecke (1990), Yager (1986), Dubois and Prade (1986), Moral (1985), Delgado and Moral (1987)). It is a ‘stochastic’ matrix\(^1\) where, for \( A, B \subseteq \Omega \), the elements \( s(B, A) \) are null if \( B \not\subseteq A \) and otherwise \( s(B, A) \) is the part of the bbb \( m(A) \) that is transferred to \( B \).

Let \( m_0 \) be Your initial bba on \( \Omega \) (represented as a row vector as every bba in this paper). We write \( m_1 = m_0 \cdot S \) to denote that \( m_1 \) results from the application of the specialization matrix \( S \) to \( m_0 \). We say then that \( m_1 \) is a specialization of \( m_0 \). \( m_1 \) can be computed by the straightforward multiplication of the vector \( m_0 \) by the matrix \( S \).

### 3. Contraction is de-specialization.

The contraction (giving up a belief) is the inverse of an expansion. It can be characterized by the matrix obtained by inverting a specialization matrix. It can indeed be defined by a matrix \( S^{-1} \) where \( S \) is a specialization matrix. \( S^{-1} \) is called a **de-specialization matrix**.

Any change of belief due to expansions and contractions, and therefore to revisions, will be represented by an updating operator. So an updating operator results from the combination of specialization and de-specialization matrices. The updating from bba \( m_1 \) to bba \( m_2 \), both defined on \( \Omega \), can be represented by a pair of specialization matrices \( S_1 \) and \( S_2 \) such that:

\[
m_2 = m_1 \cdot S_1 \cdot S_2^{-1}
\]

We say that \( m_2 \) is an updating of \( m_1 \).

We define the **Dempsterian specialization matrix** \( D_m \) related to a bba \( m \) as the specialization matrix that will update any bba \( m_1 \) into the bba that would be obtained by combining \( m \) and \( m_1 \) by Dempster’s rule of combination (\( \oplus \)):

\[
m_1 \cdot D_m = m_1 \oplus m
\]


\(^1\) This means that \( s(B, A) \geq 0 \) and \( \sum_{B \subseteq A} s(B, A) = 1 \).
Let \( m_1 \) and \( m_2 \) be two bba that obey to (1). It can be shown that one can always find two Dempsterian specialization matrices \( D_1 \) and \( D_2 \) such that:

\[
m_2 = m_1 \cdot D_1 \cdot D_2^{-1} = m_1 \cdot D_2^{-1} \cdot D_1 \tag{2}
\]

In (2), the Dempsterian specialization matrices \( D_1 \) and \( D_2 \) are those related to \( m_1 \) and \( m_2 \), respectively: \( D_1 = D_{m_1} \) and \( D_2 = D_{m_2} \).

### 4. The anatomy of the evidence.

Suppose the bba \( m_A \) is an updating of a bba \( m_0 \). What are the pieces of evidence that have induced the change of belief from \( m_0 \) to \( m_A \)? We know that there exist pairs of Dempsterian specialization matrices \( D_{1A} \) and \( D_{2A} \) such that:

\[
m_A = m_0 \cdot D_{1A} \cdot D_{2A}^{-1} \tag{3}
\]

There are many pairs of Dempsterian specialization matrices that satisfy the relation (2) but one can show that each Dempsterian specialization matrix \( D \) admits a unique representation such that

\[
D = Q \cdot \Lambda \cdot Q^{-1}
\]

where \( Q \) is a constant matrix whose elements are only 0 and 1 (it is the matrix that transforms a bba row vector into a communality function represented also as a row vector) and \( \Lambda \) is a diagonal matrix whose elements are the communalities corresponding to the bba underlying the Dempsterian specialization matrix \( D \). One can further decomposed \( \Lambda \) into a product of diagonal matrices \( \Lambda_X \), \( X \subseteq \Omega \), such that the diagonal elements of each \( \Lambda_X \) are the communalities induced by the simple support function (SSF) focused on \( X \). So in general,

\[
D = Q \cdot \prod_{X \subseteq \Omega} \Lambda_X \cdot Q^{-1}
\]

(slight adaptations must be introduced if the bba underlying \( D \) is not directly separable into SFF). The product \( D_{1A} \cdot D_{2A}^{-1} \) can then be represented as

\[
D_{1A} \cdot D_{2A}^{-1} = Q \cdot \prod_{X \subseteq \Omega} \Delta_{XA} \cdot Q^{-1}
\]

where the \( \Delta_{XA} \) are either the diagonal matrix or the inverse of the diagonal matrix induced by a SSF focused on \( X \).

The set of the matrices \( \Delta_{XA} \), for \( X \subseteq \Omega \), summarizes the impact of all the pieces of evidence involved in the updating from \( m_0 \) to \( m_A \).

### 5. Combining pieces of evidence .

Suppose the bba \( m_A \) and \( m_B \) are two updatings of a bba \( m_0 \). Let the two sets of matrices \( \Delta_{XA} \) and \( \Delta_{XB} \), \( X \subseteq \Omega \), be defined as above. Let \( m_{AB} \) be the bba that corresponds to the combination of all the pieces of evidence that have induced \( m_0 \), \( m_A \) and \( m_B \). The matrices \( \Delta_{XA} \) and \( \Delta_{XB} \) summarize the impact of the pieces of evidence that are included in \( A \) and
on B and not considered in \( m_0 \). So \( m_{AB} \) must result from all these pieces of evidence and the result can be shown to be:

\[
m_{AB} = m_0 \cdot Q \cdot \prod_{X \subseteq \Omega} \Delta_{XA} \cdot \Delta_{XB} \cdot Q^{-1}
\]

The result reduces into Dempster's rule of combination

\[
m_{AB} = m_A \oplus m_B
\]

if \( m_0 \) is a vacuous belief function.

In fact, \( m_0 \) could be seen as the 'correlation' between \( m_A \) and \( m_B \). The absence of correlation (independence) translated then into the assumption that \( m_0 \) is a vacuous bba. We use the word 'distinct' to qualify such a form of 'independence' between two pieces of evidence. (The word 'independence' is not appropriate as it describes a property between some subsets of \( \Omega \). The word 'distinctness' is more appropriate as it describes a property between two sets of pieces of evidence.) Within the TBM, the 'correlation coefficient' happens to be the whole belief function \( m_0 \).

**Definition of 'Distinct pieces of evidence'**.

Two pieces of evidence are distinct if and only if the bba common to the bba they induce is vacuous.

The problem of recognizing distinctness becomes essentially a problem of acknowledging that there is only a vacuous correlation and that both \( \Delta_{XA} \) and \( \Delta_{XB} \) results from unrelated, distinct pieces of evidence. It can not be achieved by only comparing \( m_1 \) and \( m_2 \). If one knows also the bba \( m_{AB}^* \) induced by the conjunction of the two pieces of evidence that individually induces \( m_A \) and \( m_B \), then it becomes easy to decide if the two pieces of evidence are distinct or not: compare \( m_{AB}^* \) with \( m_{AB} \). A difference reflects a non-vacuous correlation\(^1\).

The real problem appears when \( m_{AB}^* \) is unknown, and one would like to build the bba induced by the conjunction of the two pieces of evidence that have induced \( m_A \) and \( m_B \). Distinctness has to be assumed. It cannot be accepted as a default rule (as in probability theory where accepting independence must result from a voluntary act, not an act by default). Its acceptance results from an in-depth comparison of the origin of the pieces of evidence that have induced \( m_A \) and \( m_B \). It is analogous to the process used is statistics and by which we accept that two experimental results are independent.

\(^1\) The computation of the corelation is then easy: the commonality function \( q_0 \) related to \( m_0 \) is:

\[
\forall A \subseteq \Omega, q_0(A) = \frac{q_1(A)q_2(A)}{q_{AB}^*(A)}
\]

where \( q_{AB}^* \) is the commonality function related to \( m_{AB}^* \).
6. Parallelism with probability independence.

One might be tempted to disregard our definition of distinctness because it seems circular or vacuous. Therefore we now show that our definition obeys the same pattern of reasoning as the one that is being followed when defining the concept of independence in probability theory. In the following paragraphs, the $P_i$’s are propositions in probability theory, the $B_i$’s are their equivalent in the TBM.

**Probability context.**
Suppose you know the probabilities $P(A)$ and $P(B)$ of two events $A$ and $B$. What is $P(A \cap B)$?

P1. If you know the correlation between the events $A$ and $B$, you can derive $P(A \cap B)$.
P2. If you can assume that $A$ and $B$ are independent events (what means the correlation is null) you obtain $P(A \cap B) = P(A)P(B)$.
P3. If you cannot assume independence between events $A$ and $B$, and you do not know the value of the correlation between them, you can use a conservative approach and use the whole set of values for $P(A \cap B)$ that are compatible with the constraints $P(A)$ and $P(B)$.
P4. In context P3, you can apply a Principle of Minimal Entropy in order to derive a point-value for $P(A \cap B)$.

**TBM context.**
When it comes to the handling of pieces of evidence within the TBM, the reasoning becomes as follows. Suppose you know two belief functions $m_A$ and $m_B$, induced by two pieces of evidence $E_A$ and $E_B$. What is the bba $m_{AB}$ that results from the conjunction of both pieces of evidence.

B1. If you know the 'correlation' (i.e. the underlying bba $m_0$), you can derive $m_{AB}$ (as done in section 5).
B2. If you can assume that $m_A$ and $m_B$ are induced by two distinct pieces of evidence (what means that you can assume that $m_0$ is vacuous), you obtain $m_{AB} = m_A \oplus m_B$.
B3. If you cannot assume distinctness between $E_A$ and $E_B$ and you do not know the value of $m_0$, you can use a conservative approach and compute the set of bba $m_{AB}$ compatible with the constraints on $m_A$ and $m_B$ (i.e. you consider all the possible bba $m_0$ and compute $m_{AB}$ for each $m_0$, in which case you end up with a set of possible $m_{AB}$).
B4. In context B3, you can apply the Principle of Minimal Commitment (Smets (1991), Hsia (1991)) in order to derive the least committed solution for $m_{AB}$ (Kennes (1991) presents the solution of this "cautious Dempster's rule of combination" when $m_A$ and $m_B$ admit a decomposition in SSF).

In probability theory, the comparison of $P(A \cap B)$ with $P(A)P(B)$ permits to test the hypothesis of independence. Analogously the comparison of $m^*_{AB}$ with $m_A \oplus m_B$ permits to test the hypothesis of distinctness. Up to here, both problems are conceptually of the same difficulty. The nice property encountered in probability theory is that independence is equivalent to $P(A|B)$,
= P(A|B), a highly intuitive property. The analogous properties we were able to derive up to now with belief functions are unfortunately not so appealing.

**Bibliography.**


SMETS Ph. (1992) The nature of the unnormalized beliefs encountered in the transferable belief model. TR/IRIDIA/92/??