Dempster-Shafer Functions as Metalinguistic Probability Functions

Branden Fitelson

Let $\mathcal{L}_n$ be a sentential language with n atomic sentences $\{A_1, \ldots, A_n\}$. Let $S_n = \{s_1, \ldots, s_{2^n}\}$ be the set of 2^n state descriptions of $\mathcal{L}_n$, in the following, canonical lexicographical truth-table order:

A_1	A_2	$\cdots$	A_{n-1}	A_n	State Description
T	T	T	T	T	$s_1 = A_1 \mathbin{\&} A_2 \mathbin{\&} \cdots \mathbin{\&} A_{n-1} \mathbin{\&} A_n$
T	T	T	T	F	$s_1 = A_1 \mathbin{\&} A_2 \mathbin{\&} \cdots \mathbin{\&} A_{n-1} \mathbin{\&} \neg A_n$
T	T	T	F	T	$s_3 = A_1 \mathbin{\&} A_2 \mathbin{\&} \cdots \mathbin{\&} \neg A_{n-1} \mathbin{\&} A_n$
T	T	T	F	F	$s_4 = A_1 \mathbin{\&} A_2 \mathbin{\&} \cdots \mathbin{\&} \neg A_{n-1} \mathbin{\&} \neg A_n$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
F	F	F	F	F	$s_{2^n} = \neg A_1 \mathbin{\&} \neg A_2 \mathbin{\&} \cdots \mathbin{\&} \neg A_{n-1} \mathbin{\&} \neg A_n$

Also, let $\Gamma = \mathcal{P}(S_n) - \varnothing$ be the set of all non-empty subsets of state descriptions of $\mathcal{L}_n$. Note, $|\Gamma| = 2^n - 1$. And, let $\Sigma \in \Gamma$ be some (unspecified) element of Γ.

Let $\boldsymbol{\mathcal{L}}_n$ be another sentential language with 2^n atoms $\{\mathcal{A}_1, \ldots, \mathcal{A}_{2^n}\}$. And, let $\{\mathfrak{s}_1, \ldots, \mathfrak{s}_{2^{2^n}}\}$ be the 2^{2^n} state descriptions of $\boldsymbol{\mathcal{L}}_n$, again, in lexicographical truth-table order. Now, give the atomic sentences of $\boldsymbol{\mathcal{L}}_n$ the following metalinguistic (with respect to $\mathcal{L}_n$) interpretation. Interpret $\mathcal{A}_i$ as asserting that $s_i \in \Sigma$, *i.e.*, that state description s_i of $\mathcal{L}_n$ is in Σ (our unspecified subset of Γ) . Hence, each state description $\mathfrak{s}_i$ of $\boldsymbol{\mathcal{L}}_n$ gives a complete specification of the members of Σ. For instance, $\mathfrak{s}_1$ asserts that $\Sigma = \{s_1, s_2, \ldots, s_{2^n}\}$, $\mathfrak{s}_2$ asserts that $\Sigma = \{s_1, s_2, \ldots, s_{2^n-1}\}$, $\ldots$, $\mathfrak{s}_{2^{2^n}-1}$ asserts that $\Sigma = \{s_1\}$, and $\mathfrak{s}_{2^{2^n}}$ asserts that $\Sigma = \varnothing$.

Next, define a probability function $\Pr(\cdot)$ on $\boldsymbol{\mathcal{L}}_n$, in terms of basic probabilities $m(\mathfrak{s}_i) = m_i$ assigned to the state descriptions of $\boldsymbol{\mathcal{L}}_n$. Thus, for all i, $m_i \geq 0$, and $\sum_i m_i = 1$. And, imagine sampling a set Σ randomly from Γ, where the sampling process is governed by the basic probability distribution m_i. That is, for instance, the probability that $\Sigma = \{s_1\}$ is just $m_{2^{2^n}-1}$, which is the probability that $\mathfrak{s}_{2^{2^n}-1}$ is true (since $\Sigma = \{s_1\}$ is precisely what $\mathfrak{s}_{2^{2^n}-1}$ asserts!). Note: the probability that $\Sigma = \varnothing$ is *zero*, since (by assumption) Γ is the set of *non*-empty subsets of S_n. Hence, $m_{2^{2^n}} = 0$. Finally, such metalinguistic probabilistic random sampling models can be used to represent *arbitrary* Dempster-Shafer functions $Bel(\cdot)$ over $\mathcal{L}_n$. Let $\bigvee \Sigma$ be the disjunction of the elements of Σ. Then,

(1) For all sentences p of $\mathcal{L}_n$, $Bel(p) = \Pr\left(\bigvee \Sigma \vDash p\right)$.[1]

Where the m_i that define our metalinguistic probability function $\Pr(\cdot)$ are just the "basic probabilities" (or "mass terms") in Dempster's representation of $Bel(\cdot)$. This shows that Dempster-Shafer functions can be represented not only as *sets of* (*object*-lingusitic) probability functions over $\mathcal{L}_n$, but also as *individual* (*meta*-linguistic) probability functions over $\boldsymbol{\mathcal{L}}_n$. And, this representation is constructive and unique. For any language $\mathcal{L}_n$ and any Dempster-Shafer function $Bel(\cdot)$ on $\mathcal{L}_n$, we can construct *the* corresponding "metalanguage" $\boldsymbol{\mathcal{L}}_n$ and probability function $\Pr(\cdot)$ over $\boldsymbol{\mathcal{L}}_n$ (plus Σ-sampling scheme) such that (1) holds.

[1] It helps to expand the right-hand side of (1), as: $\Pr(\bigvee \Sigma \vDash p) = \sum_i \Pr(\bigvee \Sigma \vDash p \mid \mathfrak{s}_i) \cdot m_i$, where $i \in [1, 2^{2^n} - 1]$. In other words, $\Pr\left(\bigvee \Sigma \vDash p\right) = \sum_j m_j$, where $j \in [1, 2^{2^n} - 1]$ and $\mathfrak{s}_j \Longrightarrow \bigvee \Sigma \vDash p$. Here, it is assumed that $\Pr(\cdot)$ assigns probability one (zero) to all metalogical truths (falsehoods).