# Belief Functions and Möbius Transforms

Kevin Kirby
Northern Kentucky University
February 20, 2005

Presentations of the Dempster-Shafer model, from Shafer's 1976 book *The Mathematical Theory of Evidence*, to expositions in Joseph Halpern's 2003 *Reasoning About Uncertainty*, deal with two mathematical objects: belief functions and mass functions. This note investigates how they can be understood in a broader setting.

We start with some provocative notation. It is nonstandard, invented for the occasion.

Let W be a finite set, and consider an arbitrary function $m$: $2^W \to [0,1]$. For an arbitrary subset $U \subseteq W$, define

$$\int_{\varnothing}^{U} m = \sum_{X \subseteq U} m(X) \qquad (1)$$

This is a sum over a lattice. Figure 1 shows the Boolean lattice of subsets of W={w,x,y,z}, with a particular function $m$ defined on it. The values of $m$ are shown in little rectangles above each point in the lattice. Let $U$ = {w,y,z}. Then this "integral" sums $m$ across the lattice from ∅ to U:

$$\int_{\varnothing}^{U} m = m(\varnothing) + m(\{w\}) + m(\{y\}) + m(\{z\}) + m(\{w,y\}) + m(\{w,z\}) + m(\{y,z\}) + m(\{w,y,z\})$$

$$= 0 + 0.05 + 0 + 0 + 0 + 0.05 + 0.05 + 0.25 = 0.40.$$

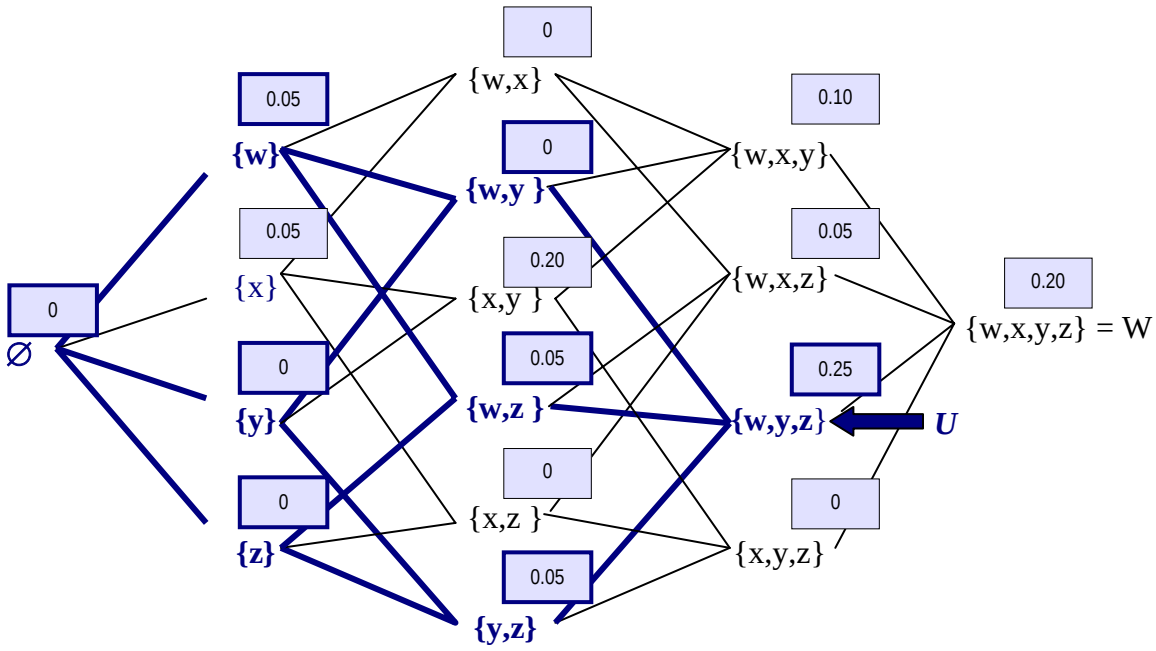



**Figure 1.** "Integrating" a function $m$ over a Boolean lattice from ∅ to $U$.

The integral notation is meant to be suggestive, of course. Notice that the particular *m* function chosen for Figure 1 has $m(\varnothing)=0$ and happens to sum to 1 across the entire lattice:

$$\int_{\varnothing}^{W} m \; = 1$$

This makes it a "mass function" in the Dempster-Shafer sense.

The restriction to an algebra of *subsets* is not necessary for the definition (1). More generally, one could work with *any* finite ordered set $(\wp, \leq)$ with a function *m* taking elements of $\wp$ to some ring, say. The idea is simple: you add up the value of *m* on all elements in the set that precede (in the ordering) a point *u*.

Recall that in elementary probability theory one has a function *f* (called a *density function* on continuous spaces, and a *mass function* on discrete spaces) and integrates it over values less than a real number *u* to get another function, the cumulative probability distribution. Supposing *f* is defined only on non-negative numbers, one has:

$$F(u) = \int_{0}^{u} f$$

Of course, this is usually done to describe a random variable. The analogy here would be with the case of a random variable which just gives the probability of a basic event. In any case, one notices immediately that $f = dF/du$ will invert this operation.

What we have done above is generalize this kind of sum from a *linearly ordered set* (like the real numbers) to a *partially ordered set* (specifically, a Boolean algebra). Pursuing the analogy, we can define the analog of a cumulative probability function on $2^W$ this way:

$$\mathrm{B}(U) = \int_{\varnothing}^{U} m$$

If *m* is a mass function, B is called a *belief* function. See Figure 2. It is amusing to note that the "thickness" of a Boolean lattice (the vertical dimension in Figure 1), varies as $\binom{|W|}{k}$, for $k$=0 to $|W|$ moving left to right, which in the limit of large $|W|$ gives us a Gaussian curve: hence the shape of the cartoon in Figure 2!

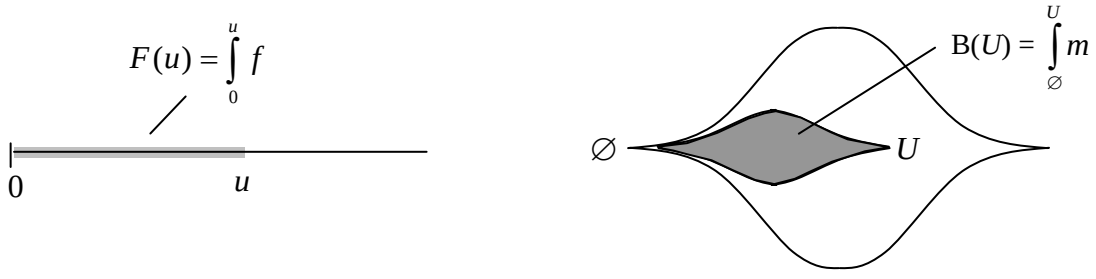



**Figure 2.** Integrating mass functions on a line and on a lattice.

Suppose we have $m$: $2^W \rightarrow [0,1]$. Given that $B(U) = \int_{\varnothing}^{U} m$ , how does one invert this and compute $m$ from B? That is, what is the Boolean lattice analogue of the differentiation of a cumulative probability distribution $F$ to get a probability density $f=dF/du$?

Let's take the set $U=\{w,y,z\}$ from Figure 1. We know $B(U)$ is the *total* $m$-value of the subsets:

$$B(U) = m(\varnothing) + m(\{w\}) + m(\{y\}) + m(\{z\}) + m(\{w,y\}) + m(\{w,z\}) + m(\{y,z\}) + m(\{w,y,z\})$$

That is,

$$m(U) = B(U) - [\, m(\varnothing) + m(\{w\}) + m(\{y\}) + m(\{z\}) + m(\{w,y\}) + m(\{w,z\}) + m(\{y,z\})\, ]. \quad (2)$$

That was easy! This is a perfectly fine recursive definition of $m$ in terms of B. In general terms:

$$m(U) = B(U) - \sum_{V \subset U} m(V)$$

where $\subset$ denotes *strict* subset. The $m$-value at $U$ is simply the *accumulated* value there (given by B), minus the total $m$-value of all the strict predecessors of $U$.

Still, a closed form expression might be nicer, so we should try to eliminate all those $m$'s on the right hand side.

Returning to the specific example in (2), let's eliminate those $m$'s one by one, working from the largest subsets to the smallest. Starting with the subsets of size two:

$$\begin{aligned}
m(\{y,z\}) &= B(\{y,z\}) - [m(\varnothing) + m(\{y\}) + m(\{z\})\,] \\
m(\{w,z\}) &= B(\{w,z\}) - [m(\varnothing) + m(\{w\}) + m(\{z\})\,] \\
m(\{w,y\}) &= B(\{w,y\}) - [m(\varnothing) + m(\{w\}) + m(\{y\})\,]
\end{aligned}$$

Turning to the subsets of size one:

$$\begin{aligned}
m(\{w\}) &= B(\{w\}) - [m(\varnothing)] \\
m(\{y\}) &= B(\{y\}) - [m(\varnothing)] \\
m(\{z\}) &= B(\{z\}) - [m(\varnothing)]
\end{aligned}$$

Finally, for the empty set, we must have $m(\varnothing) = B(\varnothing)$.

We can then substitute these into each other and put everything back into (2), which, after some simplification, gives us an expression solely in terms of B values:

$$\begin{aligned}
m(U) = \;& B(U) \\
& - [B(\{y,z\}) + B(\{w,z\}) + B(\{w,y\})] \\
& + [\, B(\{w\}) + B(\{y\}) + B(\{z\})\,] \\
& - [B(\varnothing)]
\end{aligned}$$

This is how we get $m(U)$ in terms of $B(U)$. Notice that we have grouped the terms by sizes of subsets, and notice the alternating signs. The example was fairly generic, and suggests that by doing some induction we will get a formula for a general $U$:

$m(U)$ = total B value of all subsets of size $|U|$
– total B value of all subsets of size $|U| - 1$
+ total B value of all subsets of size $|U| - 2$
– total B value of all subsets of size $|U| - 3$
+ … etc., down to B(∅).

This is quite simple, although this simplicity is obscured somewhat when we write it down as a compact formula:

$$m(U) = \sum_{i=0}^{|U|} (-1)^{|U|-i} \sum_{\substack{X \subseteq U \\ |X|=i}} B(X) \tag{3}$$

This is the someone messy analogue of $f=dF/du$. But it clearly resembles a derivative: we are extracting the contribution of a point $U$ to an integral. It is the inverse operation we were seeking. So, why not, let's use the notation $m = d\mathrm{B}$.

A final mathematical word. All of the above mathematics could be done on structures that are slightly more general. It looks like $\int$ and $d$ could be defined on *any* partially order set $(\wp, \leq)$ that has the property that if $X \leq Y$, the set $\{Z \mid X \leq Z \leq Y\}$ is finite. This kind of *local finiteness* still allows $\wp$ to be infinite. The Boolean algebra $\wp = 2^X$ for a finite set X is a special case, with the role of ≤ played by ⊆. An integral can be defined between any two comparable elements:

$$\int_A^B m = \sum_{A \leq X \leq B} m(X)$$

What definition could be more natural? But with a partial order, some elements are not comparable. For example, if A and C are not related by ≤, then the integral from A to C is undefined. This is shown in Figure 3.

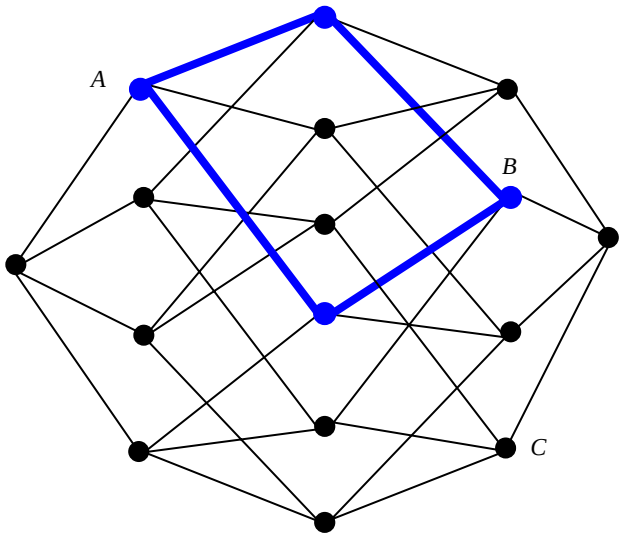



**Figure 3.** $\int_A^B m$ is defined on this partial order and the points over which $m$ is summed are shown in blue. The integral $\int_A^C m$ is undefined.

The definition of the "derivative" of $B$ in (3) above would need some rephrasing so as not to depend specifically on the algebra of subsets; the cardinality operation $|U|$ is out of place in the general context. In general, the alternating signs just serve to reduce multiple counting of points on the lattice. One could write a general graph-based algorithm for doing this calculation, but a specific formula like (3) would depend on the specific model of the abstract lattice. In general, the lattice will not come in nice "layers" (the fibers of the cardinality map, for example). A general formula could be set up like this:

$$\mathrm{d}B(\mathrm{U}) = \sum_{X \le U} K(U, X) B(X)$$

Where $K(U,X)$ (a kind of kernel) depends on the specific nature of the partial order. In the boolean algebra of subsets case (3), one takes

$$K(U,X) = (-1)^{|\mathrm{U}|-|\mathrm{X}|}. \tag{4}$$

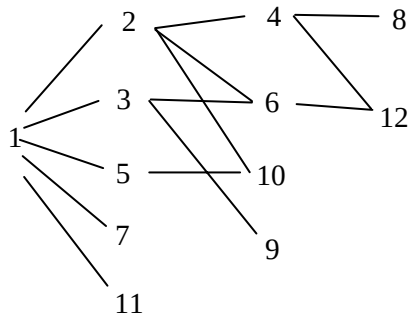
Interestingly, there is another concrete instantiation of this $\int$ and $d$ business besides the boolean algebra of subsets. This is hinted at by the use the term "Möbius transform" for the $m \leftrightarrow B$ relationship used in Shafer's book. This is clearly not related to the Möbius transformation of complex analysis; rather it seems to be related to something in number theory.

In D.M. Brown *Elementary Number Theory* (3d Ed, 1994), pp. 112-117 we find the following:

> **Theorem** (Möbius Inversion Formula). Let $F$ and $f$ be two number-theoretic functions related by the formula
>
> $$F(n) = \sum_{d|n} f(d)$$
>
> Then
>
> $$f(n) = \sum_{d|n} \mu(n/d) F(d)$$
>
> where $\mu(n)$ is the Möbius function.

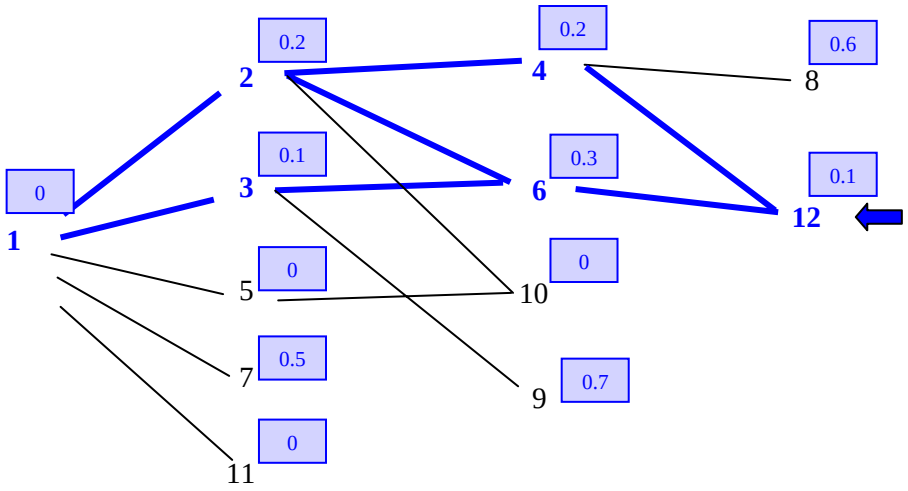
The *Möbius function* gives information about prime factorizations. $\mu(n)$ is defined as zero unless $n$ is a product of $r$ distinct primes; if it is such a product, then the value of $\mu(n)$ is $(-1)^r$. For example, $\mu(6)= \mu(2\times3)= 1$, $\mu(30)=\mu(2\times3\times5)= -1$, $\mu(12)=\mu(2^2\times3)= 0$.

What is the connection to Dempster-Shafer? I suspect it is this. The "divides" relation ("$n|m$") is reflexive, antisymmetric and transitive, so it defines a partial order on the positive integers. Even though this is an infinite set, it is "locally finite" in the sense above. A fragment of this infinite lattice looks like this:



The "|" relation is the transitive closure of the relation given by the arcs in this figure. I have tried to draw it so that the arcs are the smallest set with this transitive closure. The *gcd* and *lcm* play the roles of $\cap$ and $\cup$, respectively.

Let's work an example to make the analogy to mass/belief functions clearer. Consider a function $f$ defined on the lattice of positive integers given by the blue boxes below, and consider the calculation at the point $u$=12:




Taking $F = \int f$, i.e. $F(n) = \Sigma_{d|n}\, f(n)$, we have F(0)=0, F(2)= 0.2, F(3)=0.1, F(4)= 0.4, F(6)=0.6., and F(12)=0.9.

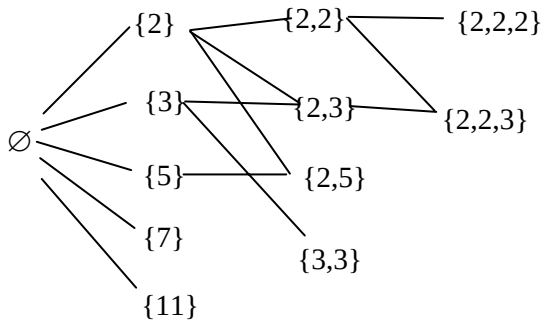
Now inverting this with $f = dF$, i.e. $\Sigma_{d|n}\, \mu\,(n/d)\, f(n)$, we can recover $f$(12):

$$
\begin{aligned}
f(12) &= \mu(12/12)F(12) + \mu(12/6)F(6) + \mu(12/4)F(4) + \mu(12/3)F(3) + \mu(12/2)F(2) + \mu(12/1)F(1) \\
&= \mu(1)F(12) \quad + \mu(2)F(6) \quad + \mu(3)F(4) \quad + \mu(4)F(3) \quad + \mu(6)F(2) \quad + \mu(12)F(1) \\
&= (1)\,(0.9) \quad + (-1)(0.6) \quad + (-1)(0.4) \quad + (0)(0.1) \quad + (1)(0.2) \quad + (0)(0) \\
&= 0.1
\end{aligned}
$$

The complicated nature of undoing "accumulation" across multiple crossing paths is common to both examples. This is what makes the interpretation of the *m* functions in Dempster-Shafer theory so strange: how can we "apportion" belief to an event and somehow bracket out all the beliefs apportioned to more specific events (subsets)?

[In Dempster-Shafer theory, beliefs B and B′ are defined by combining their derivatives *d*B and *d*B′ using Dempster's combination formula. Perhaps the "transform" language suggested by "Möbius transform" is valuable in this sense. Compare this with the Fourier transform. It moves us from an intuitive "time" domain to a "frequency" domain which, though less intuitive, is where an operation that is complicated (e.g. convolution) becomes simple (multiplication). In the same way, the Möbius transform takes us from a more intuitive setting (that of belief functions) to a less intuitive setting, but one which allows an easier formulation of belief combination.]

The number-theory application may seem to be conceptually quite different from the power set application used in Dempster-Shafer theory, but perhaps there is an even closer kinship than suggested above. Think of how the *divides* relation and the Möbius function view integers: an integer is just a *bag* of prime factors! (A *bag* is a set with repetitions allowed.) In fact, let's translate the figure above by writing integers as bags of their prime factors, order by the *sub-bag* relation:




So the essence of $\mu$ merely has to do with bags of elements. The Möbius function is zero on bags with repeated elements; it returns ±1 on bags that look like sets, i.e. have no repeated elements. Specifically, if X is a bag with no repetitions, $\mu(X)= (-1)^{|X|}$. In the transformation from number-theoretic *F* to *f*, the value $\mu(n/d)$ is used. In bag-theoretic terms, integer division is the deletion of certain prime factors from the bag, so the bag-theoretic analog needed is $\mu(U–X) = (-1)^{|U-X|} = (-1)^{|U|-|X|}$. This is the same as (4) above. In other words, the Dempster-Shafer theory uses a special case of Möbius transforms, the special case where all bags are sets.

In summary, we have the following general theory and four examples:

Let $(\wp,\leq)$ be a locally finite partial order with universal lower bound O. Given any function $m: \wp \to \mathbf{R}$, and any $U \in \wp$, define:

$$B(U) = \int_{O}^{U} m = \sum_{X \leq U} m(X)$$

The operation *d* inverts this: *d*B= *m*. The inversion formula that gives a derivative via an integral:

$$\mathrm{d}B(\mathrm{U}) = \sum_{X \leq U} K(U, X) B(X)$$

**Example 1:** $\wp = \{0,\ldots,n\}$
$\leq$ = the *less-than-or-equal* relation
$\mathrm{B}(U) = \sum_{i=0}^{U} m(X)$
$K(U,X)$ = +1 if $X=U$,
= –1 if $X=U–1$
= 0 otherwise [since $m(U) = \mathrm{B}(U) – \mathrm{B}(U–1)$].

**Example 2:** $\wp = 2^{W}$ for some finite set W.

$\leq$ = the *subset* relation

$$B(U) = \sum_{X \subseteq U} m(X)$$

$K(U,X) = (-1)^{|U|-|X|}$.

**Example 3:** $\wp = \mathbf{Z}^{+}$

$\leq$ = the *divides* relation

$$B(U) = \sum_{X | U} m(X)$$

$K(U,X) = \mu(U/X)$, where $\mu$ is the number-theoretic Möbius function.

**Example 4:** $\wp$ = AllBags[W] for some finite set W; this set is countably infinite

$\leq$ = the *subbag* relation

$$B(U) = \sum_{X \subseteq U} m(X)$$

$K(U,X) = \mu(U-X)$, where $\mu$ is the bag-theoretic Möbius function and "–" is bag-theoretic difference

As an amusing endnote, the idea of computing a derivative using an integral is not unfamiliar. One recalls this little piece of classic complex analysis:

$$dB(u) = \oint K(u,x)B(x)\,dx \quad \text{where} \quad K(u,x) = \left[2\pi i(u-x)\right]^{-2}$$

Addendum

Some treatment of Möbius functions can be found in M. Hall's 1967 textbook *Combinatorial Theory* (pp. 8-18). Although the cases of sets and bags are not mentioned, the number-theoretic definition is presented and generalized to a locally finite partially ordered set.

One fascinating result (due to G.C. Rota in 1964) proved by Hall:

Let $(\wp,\leq)$ be a locally finite partial order. Consider the set of functions $\wp\times\wp\rightarrow\mathbf{R}$. This is an algebra with scalar multiplication, pointwise addition, and a multiplication of functions $h = f*g$ defined this way:

$$h(x,y) = \sum_{x\leq z\leq y} f(x,z)g(z,y)$$

Let $\zeta$ be the characteristic function of the order relation:

$$\begin{aligned}\zeta(x,y) &= 1 \text{ if } x\leq y\\ &= 0 \text{ otherwise}\end{aligned}$$

The *Möbius function* $\mu$ of a partially order set $(\wp,\leq)$ is the inverse of its characteristic function:

$$\delta = \zeta*\mu = \mu*\zeta$$

where $\delta$ is the Dirac delta (the characteristic function of =).