Probabilities, Uncertain Reasoning, and Möbius Transforms

Kevin Kirby

Northern Kentucky University

KYMAA Annual Meeting April 1-2, 2005

Degrees of Belief:
Are They Like Probabilities?

Maybe....

Bel([*It will rain today*]) = 0.7
Bel([*It will not rain today*]) = 0.3

Bel([*Jo is sad or sleepy*]) = Bel([*Jo is sad*])
$\qquad\qquad\qquad\qquad$ + Bel([*Jo is sleepy*])
$\qquad\qquad\qquad\qquad$ − Bel([*Jo is sad and sleepy*])

Maybe Not: "Frank Agnosticism"

"The rules for belief functions permit us, when we have little evidence bearing on a proposition, to express frank agnosticism by according both that proposition and its negation very low degrees of belief. [...]

"The Bayesian theory, on the other hand, cannot deal so readily with the representation of ignorance, and it has often been criticized on this account. The basic difficulty is that the theory cannot distinguish between lack of belief and disbelief."

- Shafer 1976 *A Mathematical Theory of Evidence*, pp.22-23.

Maybe Not: "Frank Agnosticism"

"The rules for belief functions permit us, when we have little evidence bearing on a proposition, to express frank agnosticism by according both that proposition and its negation very low degrees of belief. [...]

"The Bayesian theory, on the other hand, cannot deal so readily with the representation ... criticized on this account ... cannot distinguish b...

$\text{Bel}(\ [\textit{It will rain today}]\ ) \qquad = 0.2$

$\text{Bel}(\ [\textit{It will \underline{not} rain today}]\ ) = 0.1$

...?

- Shafer 1976 *A Ma...*

Belief Functions: The Axioms

A *belief function* on a finite set W is a function $\text{Bel}: 2^W \to [0,1]$ satisfying:

1. $\text{Bel}(\varnothing) = 0$

cf. probability measures, which need only be defined on a <u>*subalgebra of* 2^W</u>

2. $\text{Bel}(W) = 1$

3. $\text{Bel}(U \cup V) \geq \text{Bel}(U) + \text{Bel}(V)$
$- \text{Bel}(U \cap V)$

*would be "="
for probability
measures*

$\text{Bel}(U \cup V \cup W) \geq \text{Bel}(U) + \text{Bel}(V) + \text{Bel}(W)$
$- \text{Bel}(U \cap V) - \text{Bel}(U \cap W) - \text{Bel}(V \cap W)$
$+ \text{Bel}(U \cap V \cap W)$

etc.

"Frank Agnosticism" from the Axioms

Take an event and its complement: A, ¬A.
The three axioms imply:
$$1 = Bel(W) = Bel(A \cup \neg A) \geq Bel(A) + Bel(\neg A) - Bel(\varnothing) = Bel(A) + Bel(\neg A)$$

So with belief functions, you only need have: $Bel(A) + Bel(\neg A) \leq 1$.
But with probabilities, you must have: $Pr(A) + Pr(\neg A) = 1$.

I can believe in A a little bit,
without having to believe in ¬A a lot.

Mass Functions

Belief functions assign numbers to events based on the accumulated "mass of evidence" of their subsets.

$$Bel(U) = \sum_{X \subseteq U} mass(X)$$

One *might have thought* it would be fine just to assign masses to individual worlds (outcomes) and accumulate them like this:

$$Bel(U) = \sum_{x \in U} mass(x)$$

But Shafer's claim is that **evidence is provided at the level of specific *subsets* of W, not at *elements* of W**.

E.g. a sensor reading can provide nonzero evidence for {x,y} while providing zero evidence for {x}, {y}, and $\varnothing$.

Evidence Mass

$m(U)$ seems to be "the amount of belief committed to U that has not already been committed to its subsets [Halpern 2003, p36]."

E.g. W= {**hep**, **cir**, **gal**, **pan**} (mutually exclusive syndromes). Consider with a diagnostic test that "was positive 70% of the time when a patient had **hep** or **cir**."
Think of how an m function could represent this test:

$$m(\{\textbf{hep},\textbf{cir}\})= 0.7.$$

Ok. How should this constrain $m(\{\textbf{hep}\})$? Or $m(\{\textbf{hep},\textbf{cir},\textbf{gal}\})$? Can they both still be 0?
Shafer: Yes.

"Calculus"
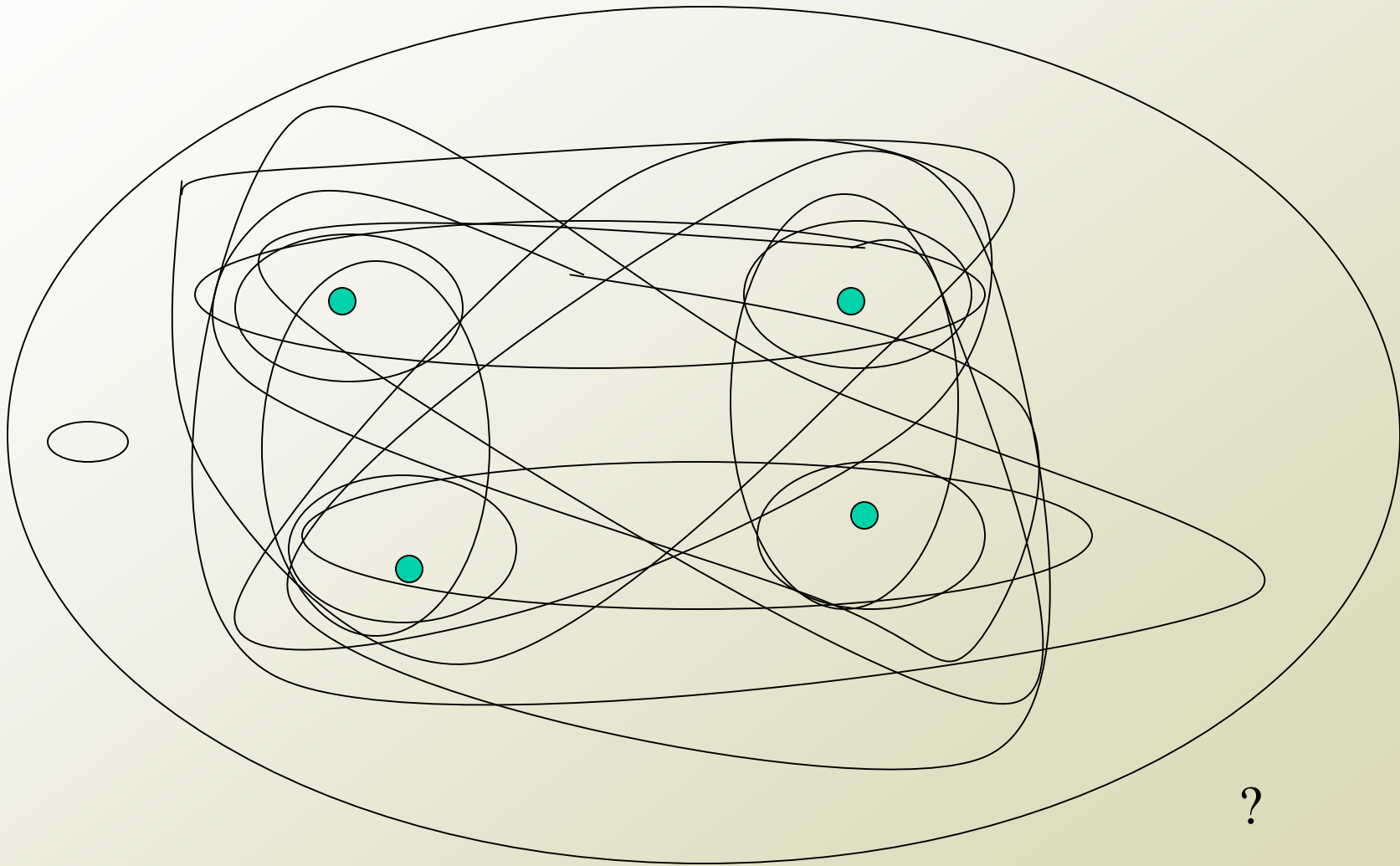
The form of a sum over all subsets of a finite set

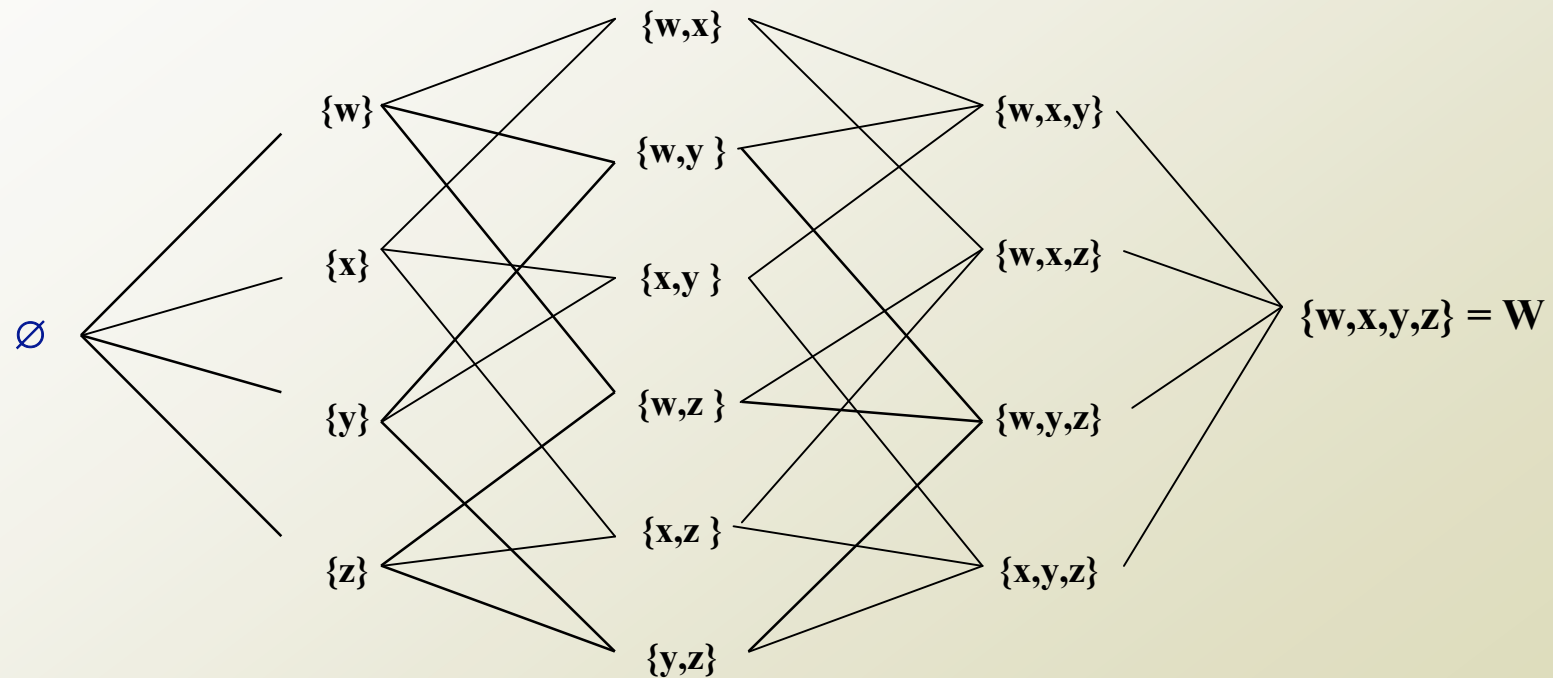$$B(U) = \sum_{X \subseteq U} m(X)$$

is a kind of **integral**.

Then: what is the related notion of **derivative**?

_____

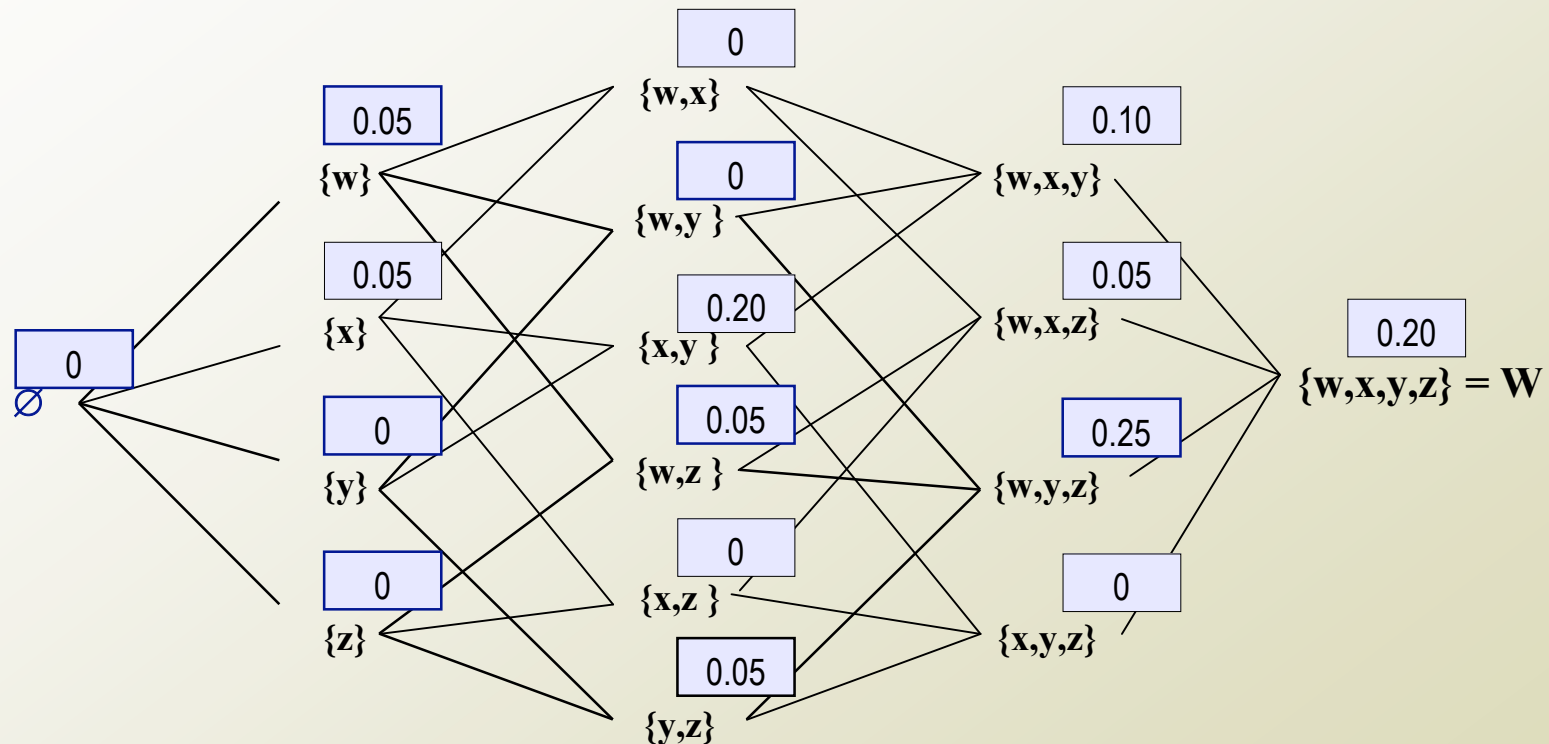To picture this, think of the **boolean lattice of subsets** of W.

Functions on Subsets

The Lattice of Subsets

A Function on the Lattice of Subsets

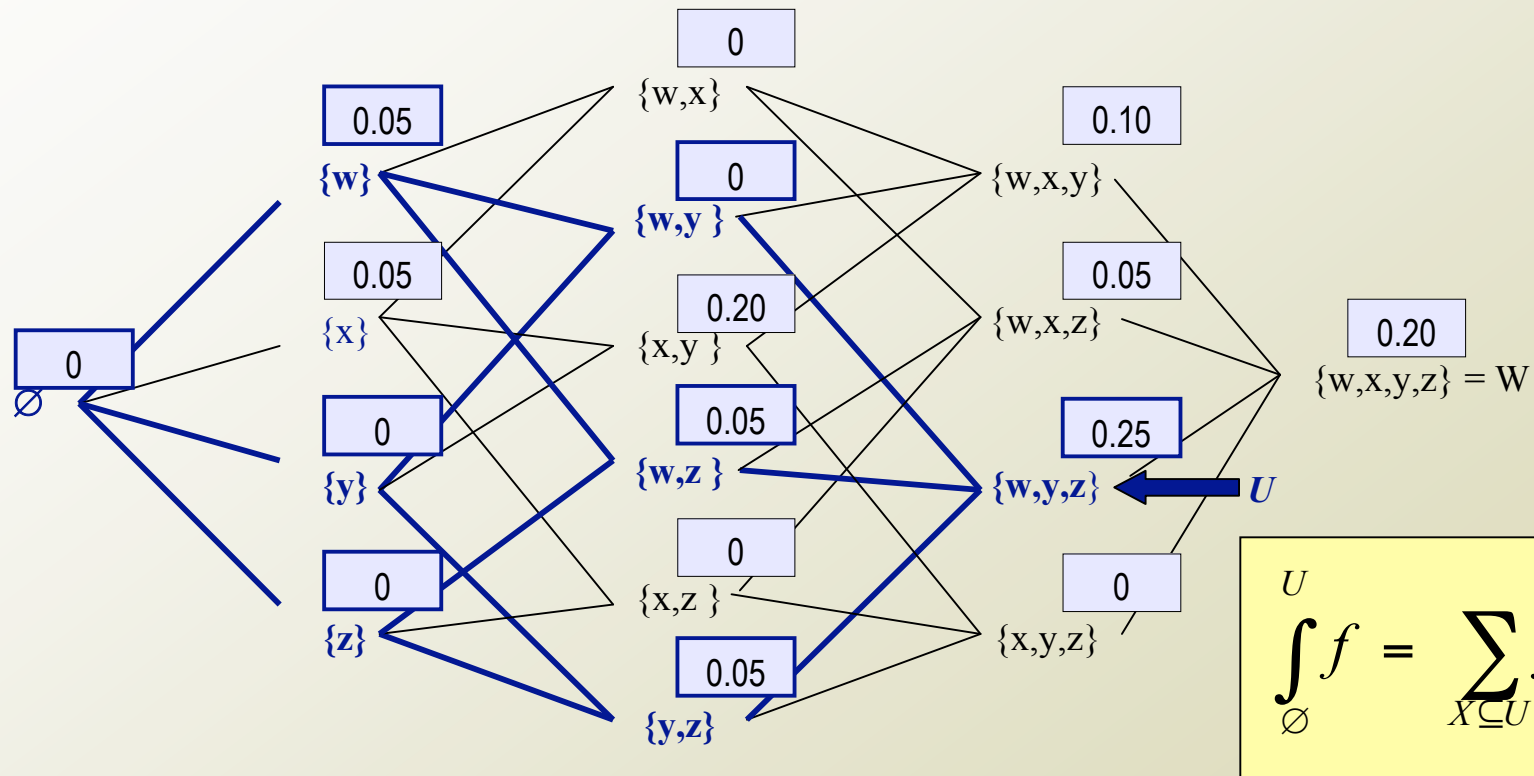
$m: 2^W \rightarrow \mathbf{R}$

Arbitrary! Here neither a probability measure nor a belief function!
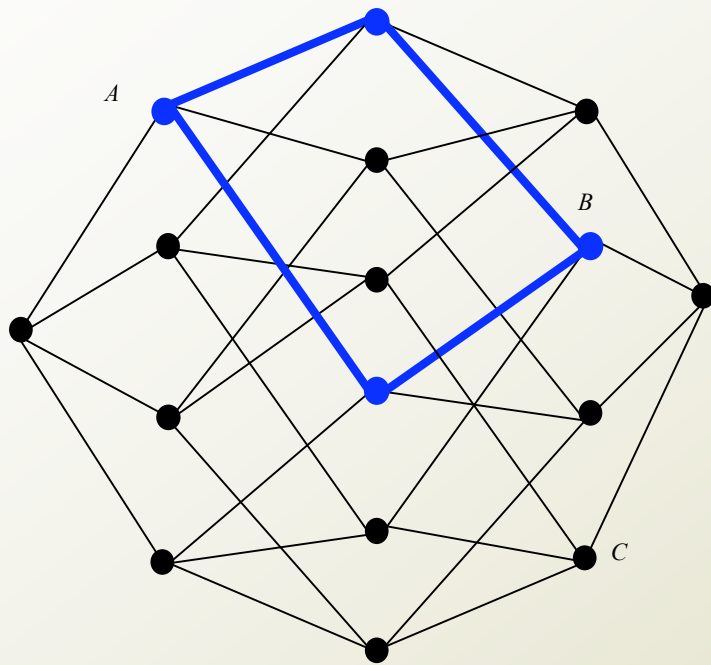
E.g. $m(\{y,z\}) = 0.05,$ $m(\{x,y,z\}) = 0$ ...

An Integral on the Lattice of Subsets

$$\int_{\varnothing}^{U} f = \sum_{X \subseteq U} f(U)$$

$$\int_{\varnothing}^{U} m = m(\varnothing) + m(\{w\}) + m(\{y\}) + m(\{z\}) + m(\{w,y\}) + m(\{w,z\}) + m(\{y,z\}) + m(\{w,y,z\})$$
$$= 0 + 0.05 + 0 + 0 + 0 + 0.05 + 0.05 + 0.25$$
$$= 0.40.$$

An Integral on Any Locally Finite Partial Order

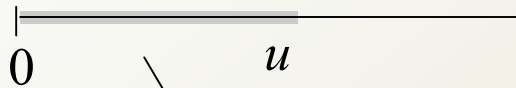

$$\int_A^B f = \sum_{A \le X \le B} f(X)$$
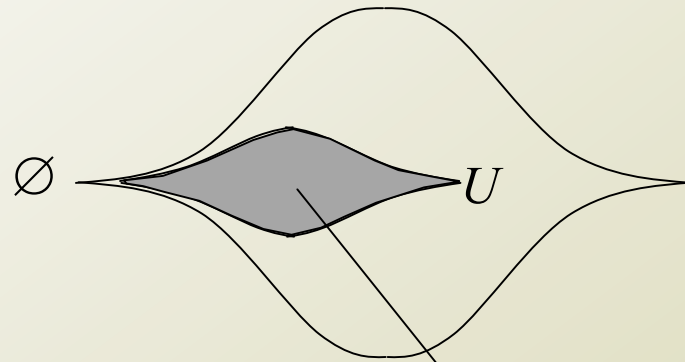
(Defined only if A and B are comparable)

"Locally finite": $\{ X \mid A \le X \le B \}$ is finite, for any A,B.

Analogies

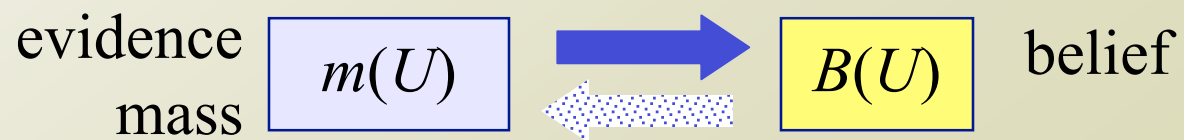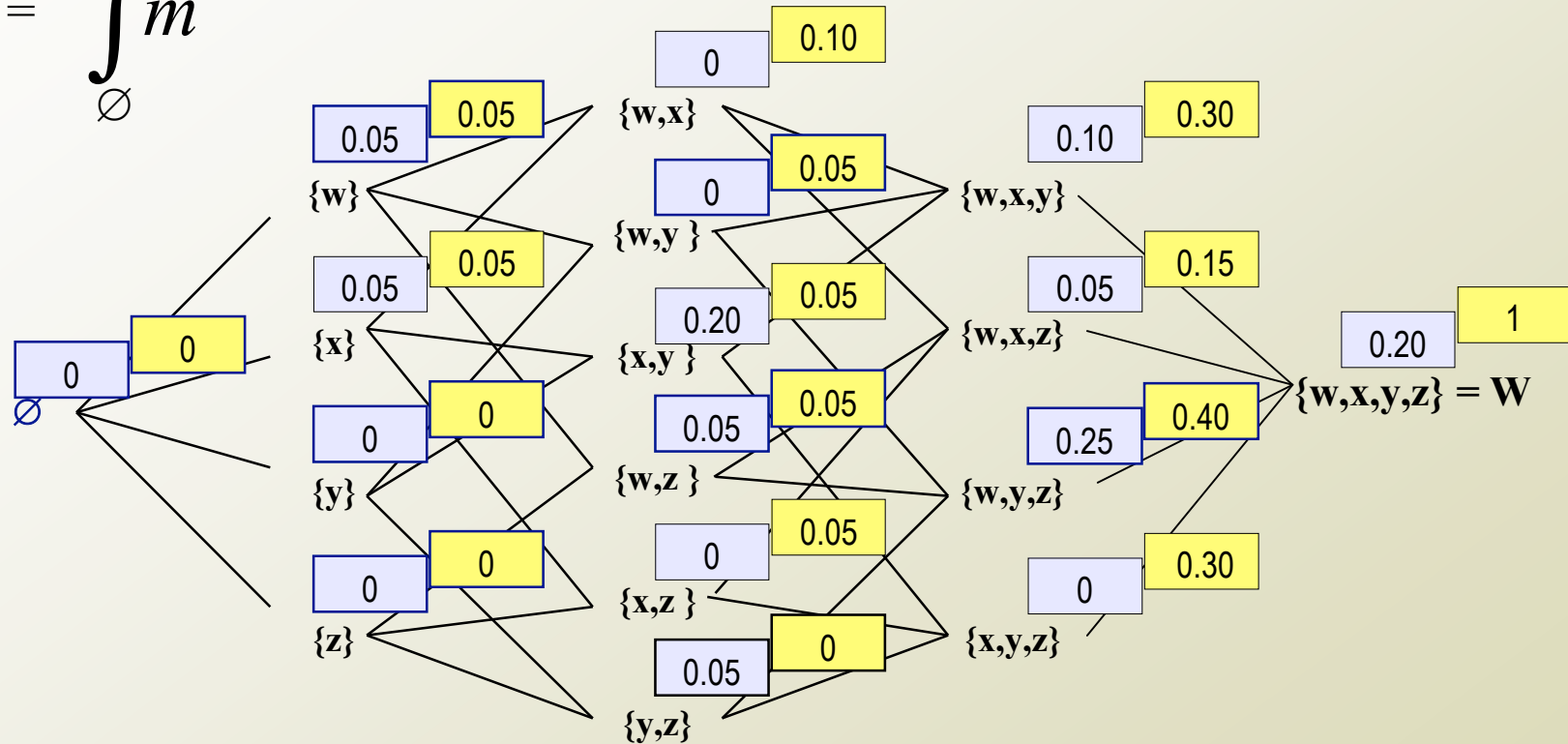$$F(u) = \int_0^u f$$

$$f = dF/du$$

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

" $m = d\mathrm{B}/dU$ " ??

Example

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

| evidence mass | $m(U)$ | $\Rightarrow$ | $B(U)$ | belief |

Getting *m* from *B*

E.g. U= {x,y,z}

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

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

This is a perfectly fine recursive definition:

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

strict subset!

Getting *m* from *B* (closed form)

Back-substitute repeatedly to eliminate this recursion:

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

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

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

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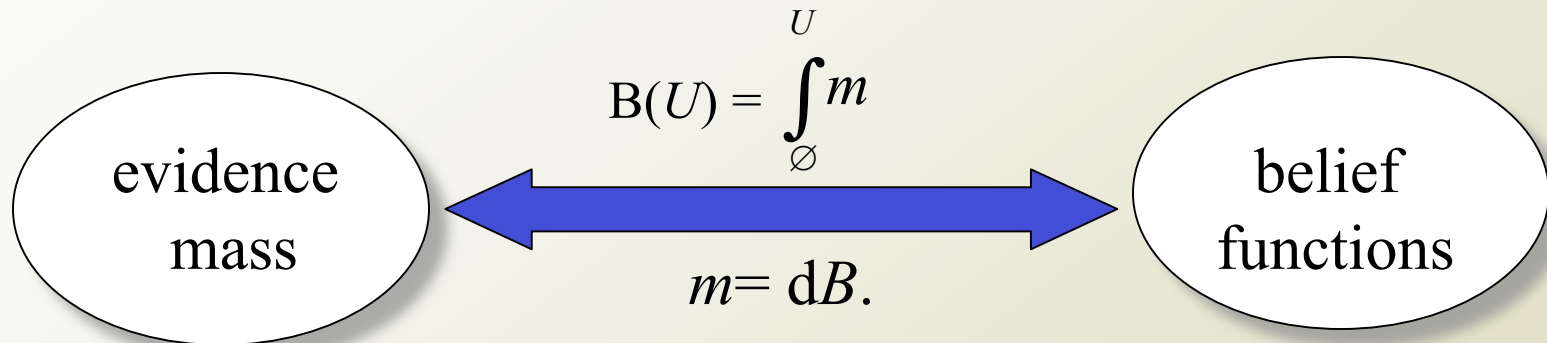$\quad\quad\quad\quad - [B(\varnothing)]$

"Derivative"

$m(U) =$ 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 $B(\varnothing)$.

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

$$= \sum_{X \subseteq U} (-1)^{|U-X|} B(X)$$

Write this $m = dB$.

Transforms

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

evidence mass ⟷ belief functions

$$m = dB.$$

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

$$m(U) = \sum_{X \subseteq U} (-1)^{|U-X|} B(X)$$

B and m are "Möbius transforms" of each other.

"Möbius?"

D.M. Brown, *Elementary Number Theory* (1994), pp.112-117:

Theorem (Möbius Inversion Formula). Let F and f be two number-theoretic functions related by the formula

$$F(n) = \sum_{d|n} f(d) \qquad \text{Then} \quad f(n) = \sum_{d|n} \mu(n/d)F(d)$$
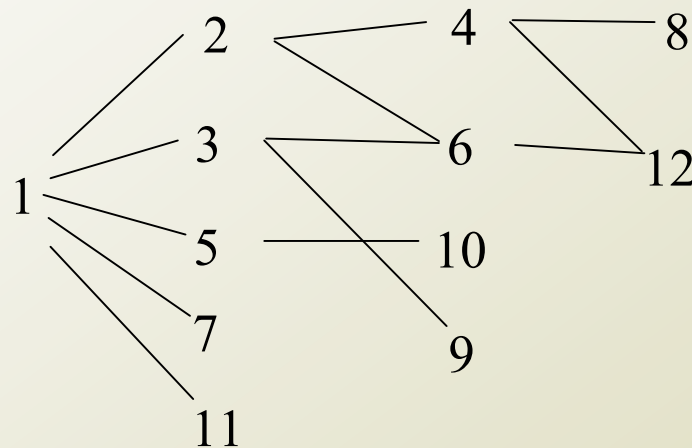
where $\mu(n)$ is the Möbius function.

The ***Möbius function*** gives information about prime factorizations.

$$\mu(n) := (-1)^r \text{ if } n \text{ is a product of } r \text{ distinct primes.}$$
$$:= 0 \qquad \text{otherwise .}$$

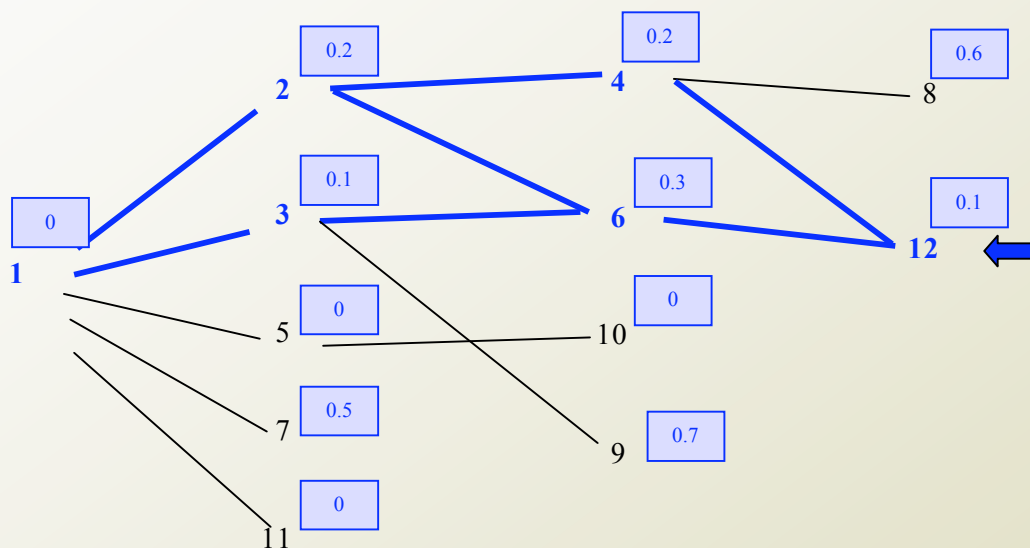$\mu(6)= \mu(2\times3)= 1 \qquad \mu(30)=\mu(2\times3\times5)= -1 \qquad \mu(12)=\mu(2^2\times3)= 0.$

Möbius and Möbius

What's the connection between this number theory application
and belief/evidence functions? The partial order.

The positive integers can be partially ordered by the *divides* relation.
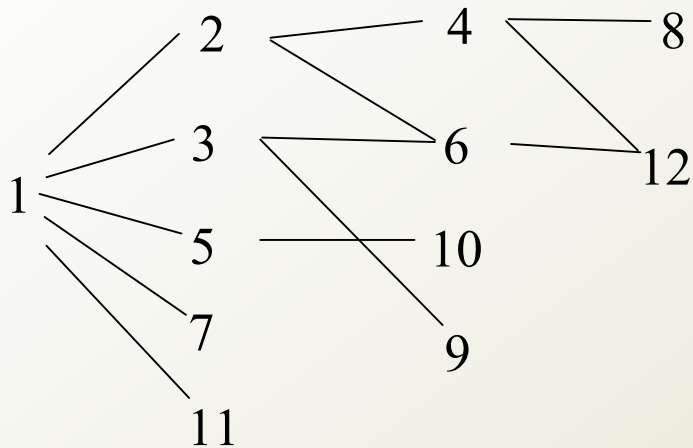LCM and GCD play the role of ∪ and ∩ in a boolean lattice.

Möbius on Integers

2 [0.2]   4 [0.2]   8 [0.6]

1 [0]   3 [0.1]   6 [0.3]   12 [0.1]

5 [0]   10 [0]

7 [0.5]   9 [0.7]

11 [0]

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.

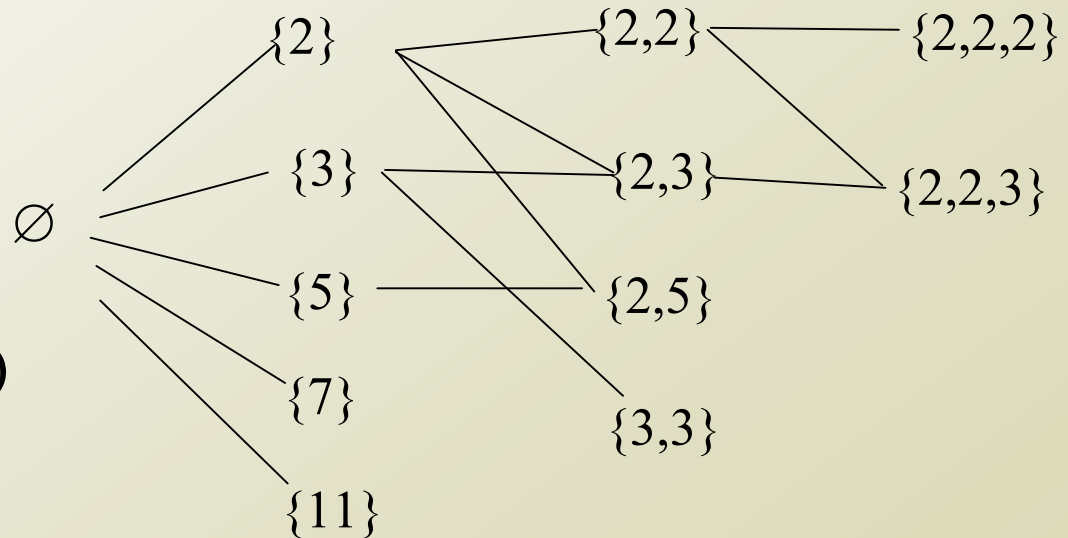Now inverting this with $f = dF$, i.e. $\Sigma_{d|n} \mu (n/d) f(n)$, we can recover $f(12)$:
$f(12) = \mu(12/12)F(12) + \mu(12/6)F(6) + \mu(12/4)F(4) + \mu(12/3)F(3) + \mu(12/2)F(2) + \mu(12/1)F(1)$
$\qquad = \mu(1)F(12) \qquad + \mu(2)F(6) \qquad + \mu(3)F(4) \qquad + \mu(4)F(3) \qquad + \mu(6)F(2) \qquad + \mu(12)F(1)$
$\qquad = (1)\,(0.9) \qquad\quad + (-1)(0.6) \qquad + (-1)(0.4) \qquad + (0)(0.1) \qquad + (1)(0.2) \qquad + (0)(0)$
$\qquad = 0.1$

The Same? Or Different?

Consider numbers merely as **bags** of their prime factors!

The same!

$$m(U) = \sum_{X \subseteq U} (-1)^{|U-X|} B(X)$$

now " − " is *bag-theoretic* difference (generalizing the set-theoretic case)

Möbius Functions, Möbius Transforms

$$m(\mathrm{U}) = \sum_{X \subseteq U} (-1)^{|U-X|} B(X)$$

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

where K(U,X) is known as a Möbius function.

Written for integers U,X instead of bags, the number theory case has K(U,X)= μ(U/X).

Think of it like a "kernel" for a Möbius transform!

Analogies Again

Getting derivatives by integrating....

$$dB(U) = \int_{\varnothing}^{U} K(U,X)B(X)\,dX$$

Reminiscent of something else....

$$dB(u) = \oint K(u,x)B(x)\,dx$$

$$K(u,x) = \left[2\pi i(u-x)\right]^{-2}$$

In General

Let $(\wp, \leq)$ be a locally finite partial order. Consider the set of functions $\wp \times \wp \rightarrow$ **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 ζ be the characteristic function of $\leq$: $\zeta(x,y) =1$ *if* $x \leq y$; $=0$ *otherwise.*

*The **Möbius function*** μ of a partially ordered set $(\wp, \leq)$ **is defined as the inverse of its characteristic function**: $\delta = \zeta*\mu = \mu*\zeta$ where δ is the Dirac delta (the characteristic function of $=$).

Rota, 1964.
See also: Aigner (1979) *Combinatorial Theory*.

Little Footnote

For more on probabilities, belief functions, etc:

J. Y. Halpern.
Reasoning Under Uncertainty.
MIT Press, 2003.

Big Footnote

Möbius functions in action!

Stephen E. Newman.
"Measure Algebras and Functions of Bounded Variation
on Idempotent Semigroups."
Transaction sof the AMS Vol. 163 (January 1972), 189-205.