A Breathless Introduction to Generalized Information Theory

Cliff Joslyn

Distributed Knowledge Systems and Modeling Team

Mail Stop B265, LANL

Los Alamos, NM 87545, USA

joslyn@lanl.gov http://www.c3.lanl.gov/~joslyn

Los Alamos NATIONAL LABORATORY

Modeling, Algorithms, and Informatics (CCS-3)

Los Alamos Workshop on

Novel Approaches to Uncertainty Quantification

February 28, 2002

ZOO OF AVAILABLE UQ FORMALISMS

Traditional:

- Set Theory
- Logic
- Probability Theory

More Novel:

- **Interval Analysis**
- **Fuzzy Systems**
- **Fuzzy and Monotone Measures**
- **Dempster-Shafer Evidence Theory**
- **Random Sets and Intervals**
- **Possibility Theory**
- **Probability Bounds**
- Rough Sets
- Imprecise Probabilities
- Probabilistic Robustness
- Info-Gap Theory

DEFER OBVIOUS CONTROVERSIES ...

- Why should I care, what's this good for, anyway?
- Isn't that just the same as probability?
- Why isn't probability good enough for you?
- Where do the numbers come from?
- Can you give me an example of where X does better than probability?
- Why do you use all those silly words like "fuzzy" and "belief"? Can you be serious?

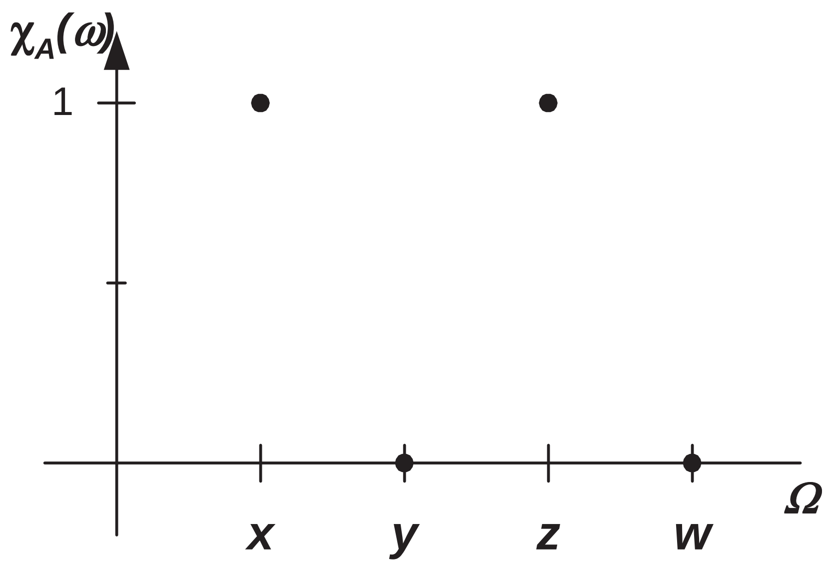
"CLASSICAL" POINT OF DEPARTURE: SETS AND LOGIC

Set Theory: $A \subseteq \Omega, \chi_A : \Omega \mapsto \{0,1\}, \chi_A(\omega \in \Omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$

Boolen Logic: $T(p) \in \{\text{false}, \text{true}\} = \{0,1\}$

Isomorphism in Boolean Algebra:

Negation	$\neg p$	$\overline{A}$
Disjunction	$p \vee q$	$A \cup B$
Conjunction	$p \wedge q$	$A \cap B$
Implication	$p \rightarrow q$	$A \subseteq B$

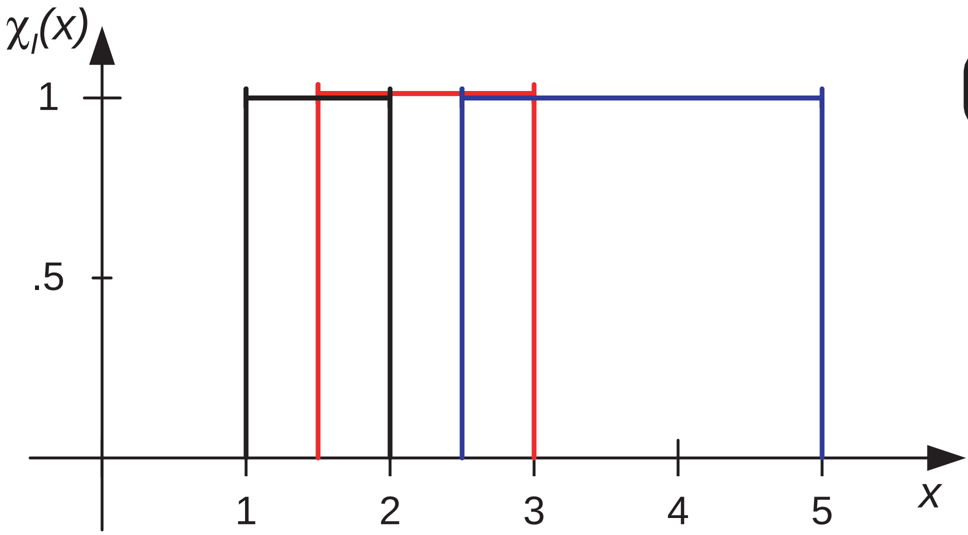


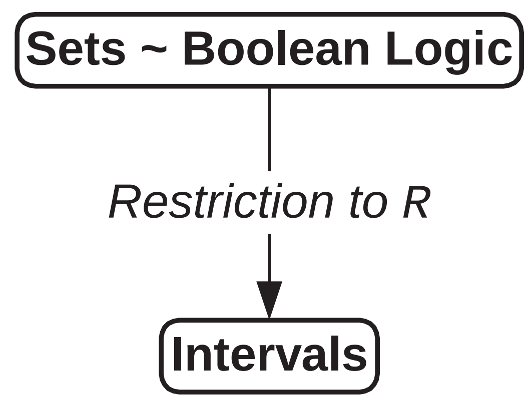
Boolen Logic = Sets

$$A = \{x, z\} \subseteq \Omega$$

INTERVALS

- $I = [I_l, I_u] \subseteq \mathbb{R}$
- $\chi_I(x) = \begin{cases} 1 & I_l \leq x \leq I_u \\ 0 & \text{otherwise} \end{cases}$
- $I_1 * I_2, * \in \{+, -, \times, \div\}$, etc.
- **Extension:** $x + y = [x, x] + [y, y]$



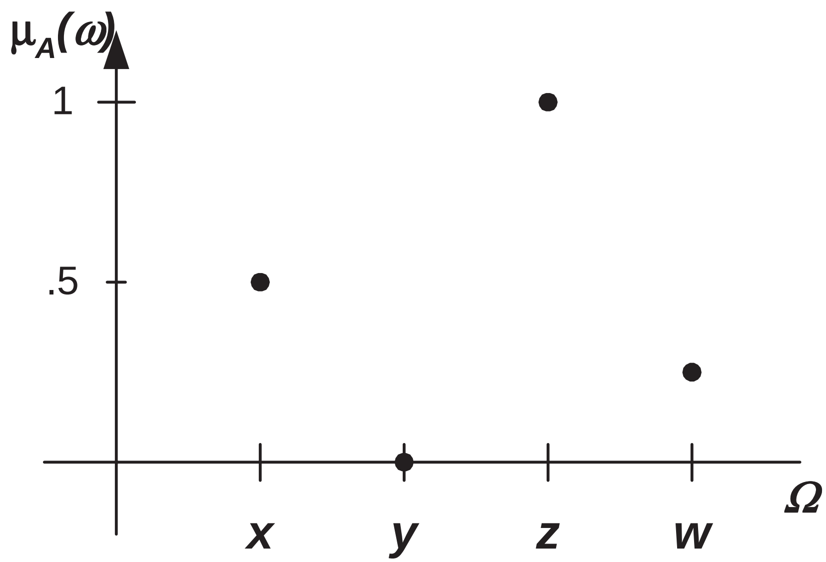


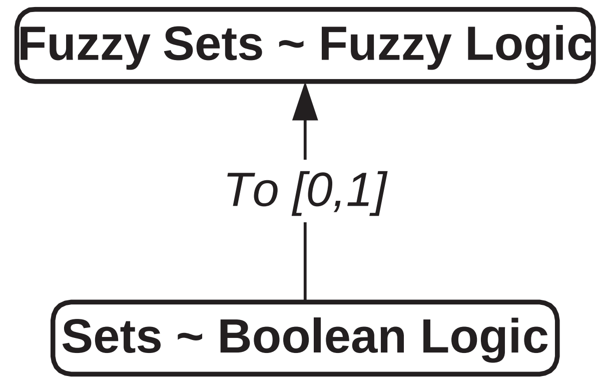
$$[1, 2] + [1.5, 3] = [2.5, 5] \subseteq \mathbb{R}$$

FUZZINESS

Fuzzy Sets: $\tilde{A} \mathrel{\tilde{\subseteq}} \Omega, \mu_{\tilde{A}} : \Omega \mapsto [0,1], \mu_{\tilde{A}}(\omega) = \tilde{A}(\omega) \in [0,1]$ is the "extent" or "degree" to which $\omega \in A$

Fuzzy Logic: $\tilde{T}(p) \in [0,1]$





$$\tilde{A} = \{\langle x, .5\rangle, \langle z, 1\rangle, \langle w, .25\rangle\} \mathrel{\tilde{\subseteq}} \Omega$$

FUZZY OPERATIONS

		(canonical example)
$\overline{\tilde{A}}$	$\mu_{\overline{\tilde{A}}}(\omega) = c(\tilde{A}(\omega))$	$\mu_{\overline{\tilde{A}}}(\omega) = 1 - \tilde{A}(\omega)$
$\tilde{A} \cup \tilde{B}$	$\mu_{\tilde{A}\cup\tilde{B}}(\omega) = \tilde{A}(\omega) \sqcup \tilde{B}(\omega)$	$\mu_{\tilde{A}\cup\tilde{B}}(\omega) = \tilde{A}(\omega) \vee \tilde{B}(\omega)$
$\tilde{A} \cap \tilde{B}$	$\mu_{\tilde{A}\cap\tilde{B}}(\omega) = \tilde{A}(\omega) \sqcap \tilde{B}(\omega)$	$\mu_{\tilde{A}\cap\tilde{B}}(\omega) = \tilde{A}(\omega) \wedge \tilde{B}(\omega)$
$\tilde{A} \subseteq \tilde{B}$	$\mu_{\tilde{A}\subseteq\tilde{B}}(\omega) = \tilde{A}(\omega) \rightarrow \tilde{B}(\omega)$	$\mu_{\tilde{A}\subseteq\tilde{B}}(\omega) = (1 - \tilde{A}(\omega)) \vee \tilde{B}(\omega)$

- $\vee, \wedge$ are max and min
- $c\colon [0,1] \mapsto [0,1]$ is a complement function

$$c(0) = 1, c(1) = 0, x \le y \rightarrow c(x) \ge c(y)$$

- $\sqcap$ ($\sqcup$) is a triangular norm (conorm) (associative copulas/co-copula): $\sqcap, \sqcup\colon [0,1]^2 \mapsto [0,1]$, associative, monotonic,

$$0 \sqcup x = x \sqcup 0 = x, \qquad 1 \sqcap x = x \sqcap 1 = x$$

$x \sqcap y$	$x \wedge y$	$\ge$	$x \times y$	$\ge$	$0 \vee (x + y - 1)$	$\ge$	$\lfloor x \rfloor \lfloor y \rfloor$
$x \sqcup y$	$x \vee y$	$\le$	$x + y - xy$	$\le$	$1 \wedge (x + y)$	$\le$	$\lceil x \rceil \lceil y \rceil$

- **Extension:** When $\tilde{A}, \tilde{B}\colon \Omega \mapsto \{0,1\}$, then "crisp" set operations recovered

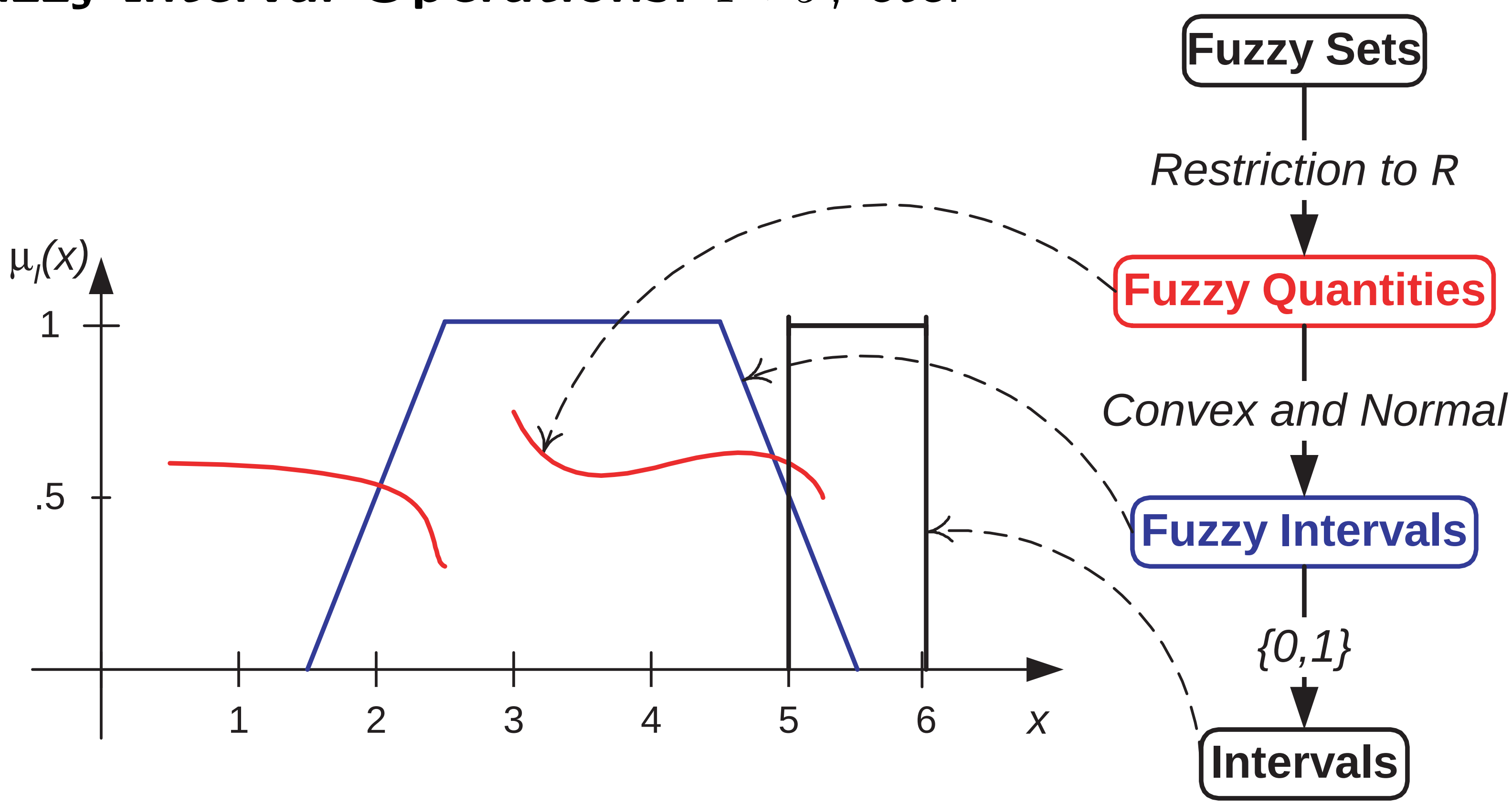
FUZZY QUANTITIES, NUMBERS AND INTERVALS

Fuzzy Quantity: Fuzzy subset of the line $\tilde{I} \tilde{\subseteq} \mathbb{R}$

Fuzzy Interval: Also convex and normal

Fuzzy Number: Single-peaked

Fuzzy Interval Operations: $\tilde{I} * \tilde{J}$, etc.



FUZZY SETS AS GENERALIZED "DENSITIES" OR "DISTRIBUTIONS"

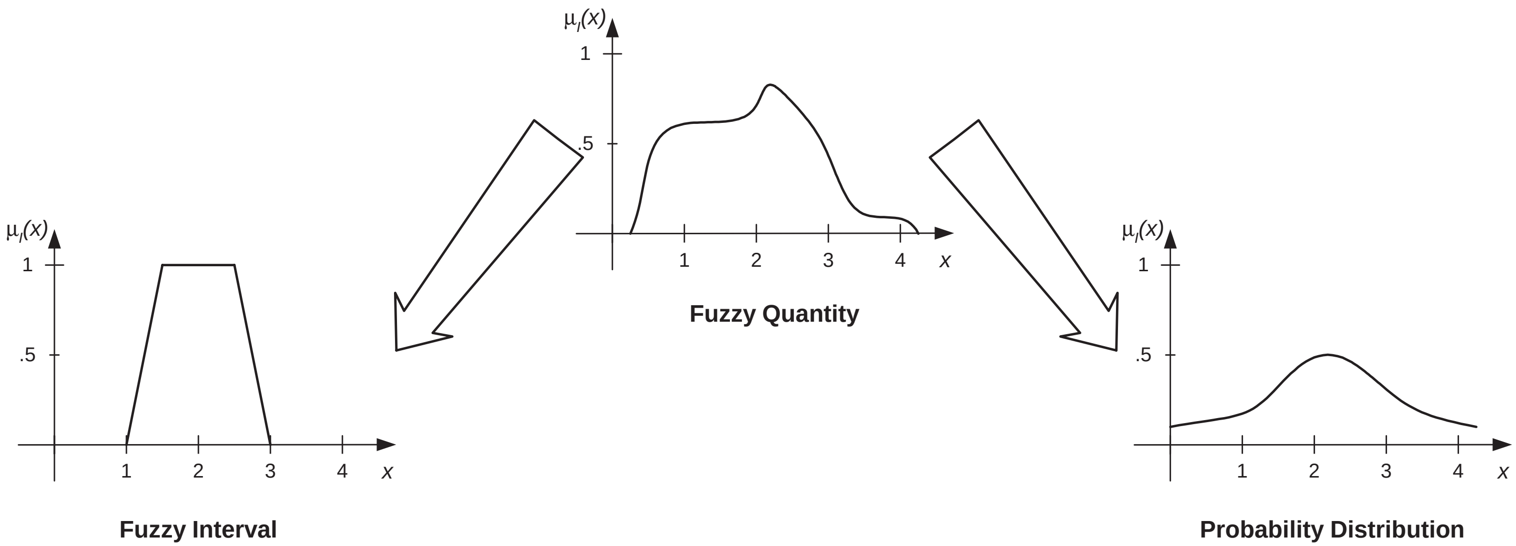
Fuzzy Quantity: $\tilde{I} \tilde{\subseteq} \mathbb{R}$

Probability Distribution: Fuzzy quantity where

$$\int_{x \in \mathbb{R}} \mu_{\tilde{I}}(x) = 1$$

Fuzzy Interval = Possibility Distribution: Fuzzy quantity where

$$\sup_{x \in \mathbb{R}} \mu_{\tilde{I}}(x) = 1$$



GENERALIZED MEASURES

- From point functions to set functions
- Generalizes relation between probability density and probability measure
- **Fuzzy Measure:** $\nu\colon 2^{\Omega} \mapsto [0,1]$, where

$$\nu(\emptyset) = 0, \quad \nu(\Omega) = 1, \quad A \subseteq B \rightarrow \nu(A) \leq \nu(B).$$

- **Trace** as concept of density $\rho_\nu\colon \Omega \mapsto [0,1], \rho_\nu(\omega) := \nu(\{\omega\})$.
- **Distributional** or **decomposable** if $\exists \sqcup, \nu(A) = \bigsqcup_{\omega \in A} \rho_\nu(\omega)$.
- **Normalization:** $\bigsqcup_{\omega \in \Omega} \rho_\nu(\omega) = 1$.
- A probability measure is a fuzzy measure ($A \subseteq B \rightarrow \text{Pr}(A) \leq \text{Pr}(B)$) with an additional additive constraint:

$$\text{Pr}(A \cup B) - \text{Pr}(A \cap B) = \text{Pr}(A) + \text{Pr}(B)$$

$$\rho_{\text{Pr}}(\omega) = p(\omega)$$

  Pr is distributional for $\sqcup = +_b$

DEMPSTER-SHAFER EVIDENCE THEORY

- **Belief** and **plausibility** as dual fuzzy measures:

$$\text{Bel}(A \cup B) \geq \text{Bel}(A) + \text{Bel}(B) - \text{Bel}(A \cap B)$$

$$\text{Pl}(A \cap B) \leq \text{Bel}(A) + \text{Bel}(B) - \text{Bel}(A \cup B)$$

$$\text{Bel}(A) + \text{Bel}(\overline{A}) \leq 1, \qquad \text{Pl}(A) + \text{Pl}(\overline{A}) \geq 1$$

$$\text{Bel}(A) = 1 - \text{Pl}(\overline{A}), \quad \text{Pl}(A) = 1 - \text{Bel}(\overline{A})$$

$$\text{Bel}(A) \leq \text{Pl}(A)$$

- Codetermined by a **basic probability assignment** $m\colon 2^{\Omega} \mapsto [0,1]$ where $\sum_{A \subseteq \Omega} m(A) = 1$:

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B), \qquad \text{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \text{Bel}(B) = \sum_{B \subseteq A} (-1)^{|A-B|} (1 - \text{Pl}(\overline{B}))$$

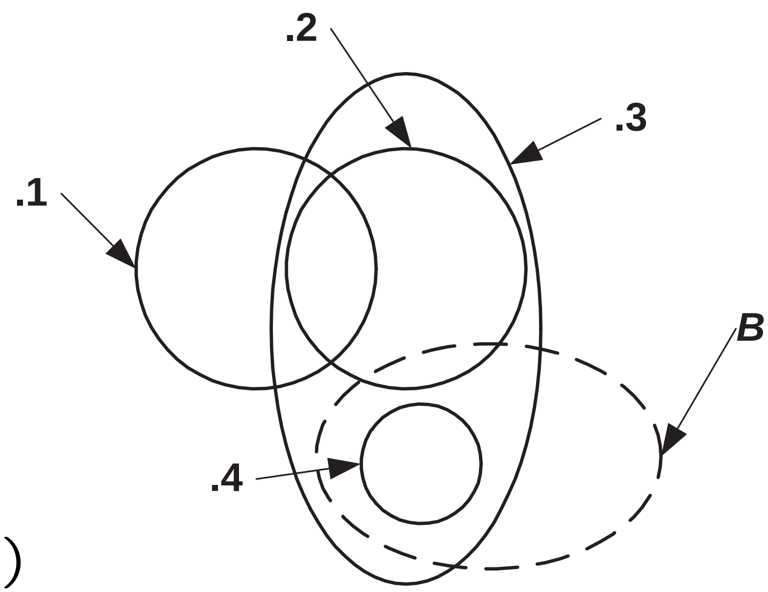
DEMPSTER-SHAFER EVIDENCE THEORY

Focal Set: $\mathcal{F} = \{A_j \subseteq \Omega : m(A_j) > 0\}$

Body of Evidence: $\langle \mathcal{F}, m \rangle = \left\langle \{A_j\}, \{m(A_j)\} \right\rangle$

Example:

$$\mathcal{F} = \{A_1, A_2, A_3, A_4\}$$
$$m(A_1) = .1, m(A_2) = .2$$
$$m(A_3) = .3, m(A_4) = .4$$
$$\text{Bel}(B) = \sum_{A_j \subseteq B} m(A_j)$$
$$= m(A_4) = .4$$
$$\text{Pl}(B) = \sum_{A_j \cap \subseteq B} m(A_j)$$
$$= m(A_2) + m(A_3) + m(A_4)$$
$$= .2 + .3 + .4 = .9$$



RANDOM VARIABLES, RANDOM SETS, AND BODIES OF EVIDENCE

- Transform a body of evidence: $\left\langle \{A_j\}, \{m(A_j)\} \right\rangle = \left\{ \left\langle A_j, m(A_j) \right\rangle \right\}$
- Recalling that $\sum_{A_j} m(A_j) = 1$, then $m(A_j)$ sure looks like a density $p(A_j)$
- **Random Variable:** Given a probability space $\langle X, \Sigma, \mathrm{Pr} \rangle$, then $S: X \mapsto \Omega$ is a random variable if S is Pr-measurable: $\forall \omega \in \Omega, \mathcal{S}^{-1}(\omega) \in \Sigma$.
- **General Random Set:** $\mathcal{S}: X \mapsto 2^\Omega - \{\emptyset\}$ is a random subset of Ω if $\mathcal{S}$ is Pr-measurable: $\forall \emptyset \neq A \subseteq \Omega, \mathcal{S}^{-1}(A) \in \Sigma$. m acts as density of $\mathcal{S}$.
- **Finite Random Set:** $\mathcal{S} = \left\{ \left\langle A_j, m(A_j) \right\rangle \right\}$.

$$\mathrm{Pr}(\mathcal{S} = A) = m(A)$$

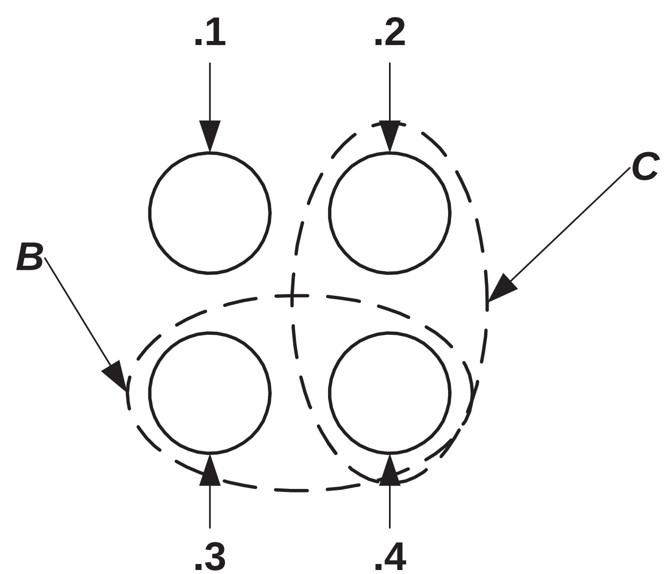
$$\mathrm{Pr}(\mathcal{S} \subseteq A) = \mathrm{Bel}(A), \qquad \mathrm{Pr}(\mathcal{S} \cap A \neq \emptyset) = \mathrm{Pl}(A)$$

DECOMPOSABLE CASE: PROBABILITY MEASURES

- Assume $\mathcal{F}$ is a disjoint class: $\forall A_1, A_2 \in \mathcal{F}, A_1 \cap A_2 = \emptyset$.
- More particularly, $\mathcal{F} = \{\{\omega\}\}$.
- $\text{Bel}(B) = \text{Pl}(B) = \text{Pr}(B)$
- Probability distribution $p(\omega) = \text{Pr}(\{\omega\}), \sqcup = +_b$:

$$\text{Pr}(A) = \sum_{\omega \in A} p(\omega), \qquad \text{Pr}(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1$$

- $\text{Bel}(B) = \text{Pl}(B) = \text{Pr}(B) = .3 + .4 = .7$
- $\text{Pr}(B \cup C) = \text{Pr}(B) + \text{Pr}(C) - \text{Pr}(B \cap C) = .9$



DECOMPOSABLE CASE: POSSIBILITY MEASURES

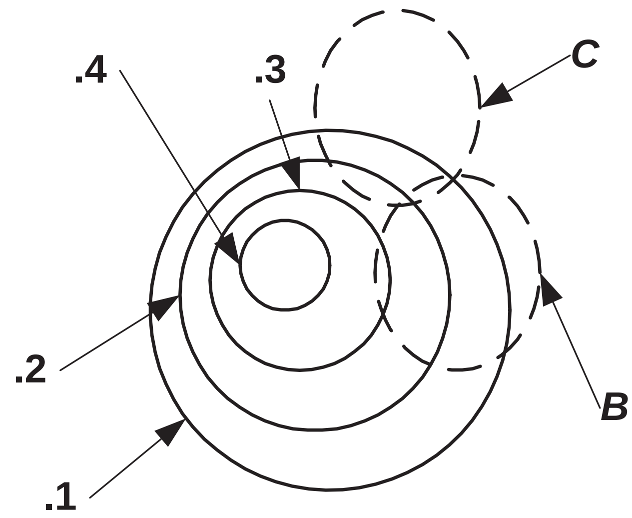
- Assume $\mathcal{F}$ is a nested class: $\forall A_1, A_2 \in \mathcal{F}, A_1 \subseteq A_2$ or $A_2 \subseteq A_1$
- Pl becomes a **possibility measure** Π distributional for $\sqcup = \vee$:

$$\Pi(B \cup C) = \Pi(B) \vee \Pi(C)$$

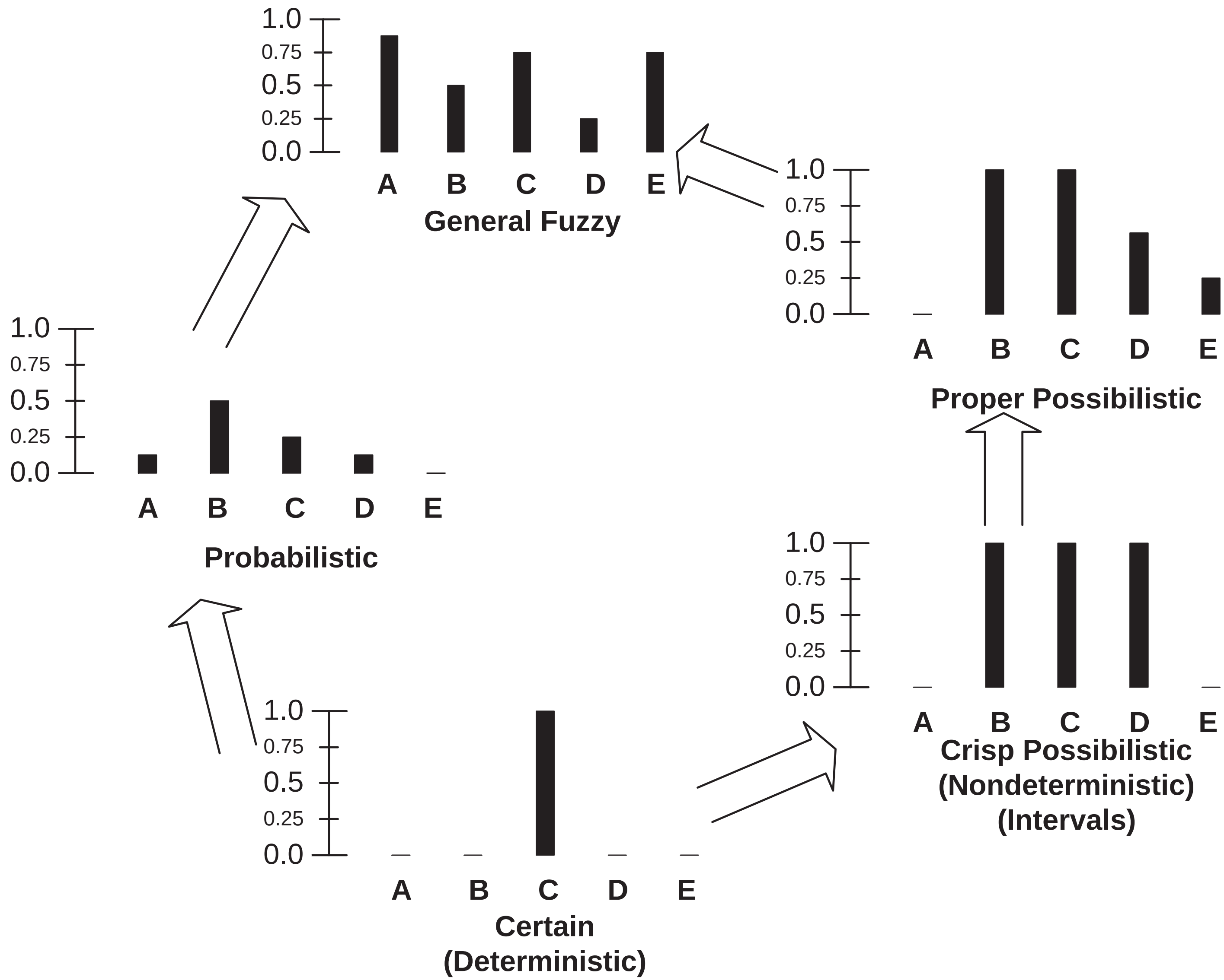
- Dual **necessity measure** $\mathrm{Bel} = \eta$
- Possibility distribution $\pi(\omega) = \Pi(\{\omega\}), \sqcup = \vee$:

$$\Pi(A) = \vee_{\omega \in A} \pi(\omega), \qquad \Pi(\Omega) = \vee_{\omega \in \Omega} \pi(\omega) = 1$$

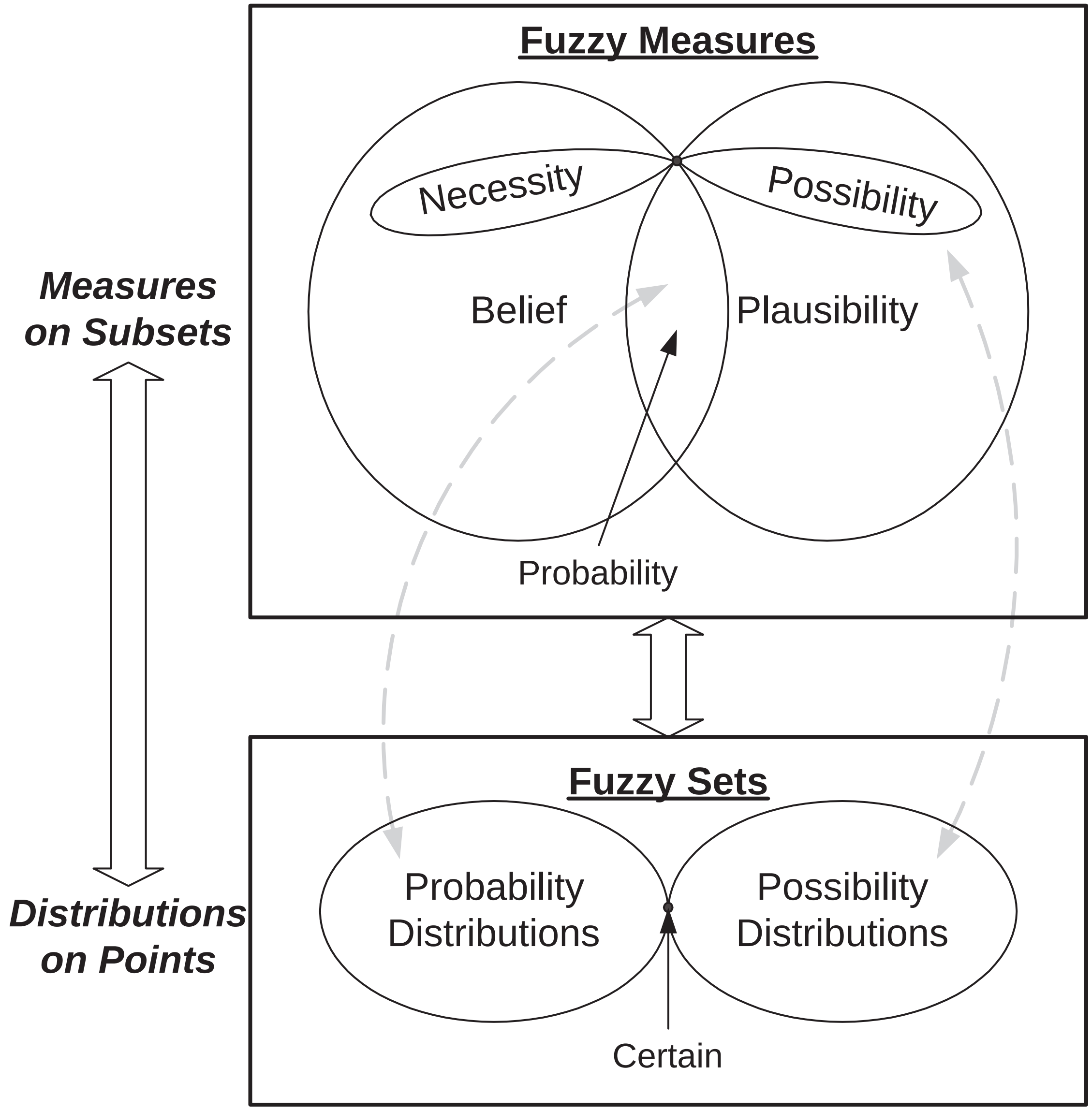
$$\begin{aligned} \Pi(B \cup C) &= \mathrm{Pl}(B \cup C) \\ &= .1 + .2 + .3 = .6 \\ &= \Pi(B) \vee \Pi(C) = .6 \vee .3 \end{aligned}$$



CLASSES OF MEASURE TRACES

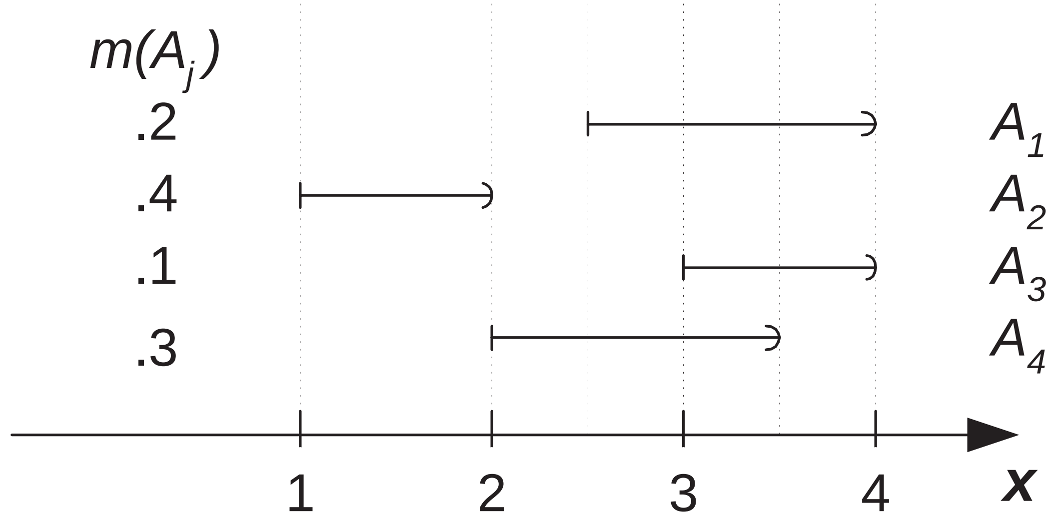


MEASURES AND DISTRIBUTIONS



RANDOM INTERVALS

- $\mathcal{D} := \{[a,b) \subseteq \mathbb{R} : a, b \in \mathbb{R}, a < b\}$.
- **Random Interval:** $\mathcal{A}$, is a random set on $\Omega = \mathbb{R}$ for which $\mathcal{F}(\mathcal{A}) = \{I_j\} \subseteq \mathcal{D}$



PROBABILITY BOUNDS

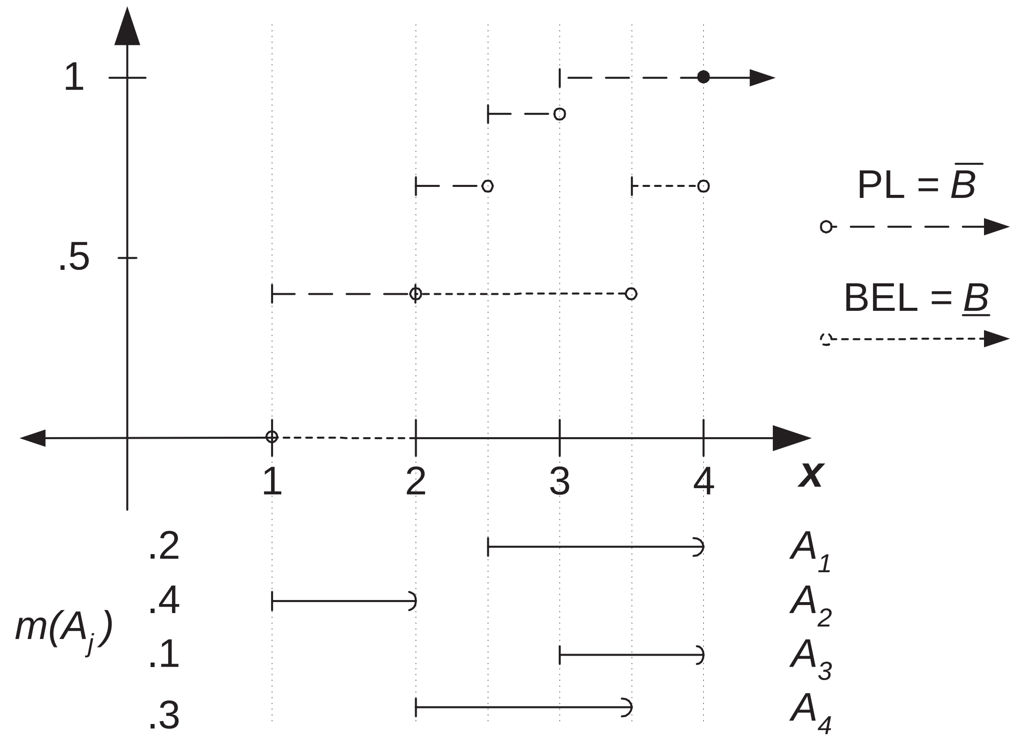
- **Cumulative Belief and Plausibility:** $\mathsf{PL}, \mathsf{BEL}\colon \mathbb{R} \mapsto [0,1]$

  $$\mathsf{BEL}(x) := \mathsf{Bel}((-\infty, x)), \qquad \mathsf{PL}(x) := \mathsf{Pl}((-\infty, x)).$$

- **Probability Bounds:** (Ferson) $\mathcal{B} := \left\langle \underline{B}, \overline{B} \right\rangle$, where $\underline{B}, \overline{B}\colon \mathbb{R} \mapsto [0,1]$, $\lim_{x \longrightarrow -\infty} B(x) \longrightarrow 0, \lim_{x \longrightarrow \infty} B(x) \longrightarrow 1$, and $\underline{B}, \overline{B}$ are monotonic with $\underline{B} \leq \overline{B}$.

RANDOM INTERVALS TO PROBABILITY BOUNDS

- Given a random interval $\mathcal{A}$, then $\langle \text{BEL}, \text{PL} \rangle$ is a PBound.
- Given a PBound $\mathcal{B}$, then $\overline{B} - \underline{B} = \rho_{\text{Pl}}$
- If a random interval $\mathcal{A}$ is consonant, then $\overline{B} - \underline{B} = \pi$ is its possibility distribution

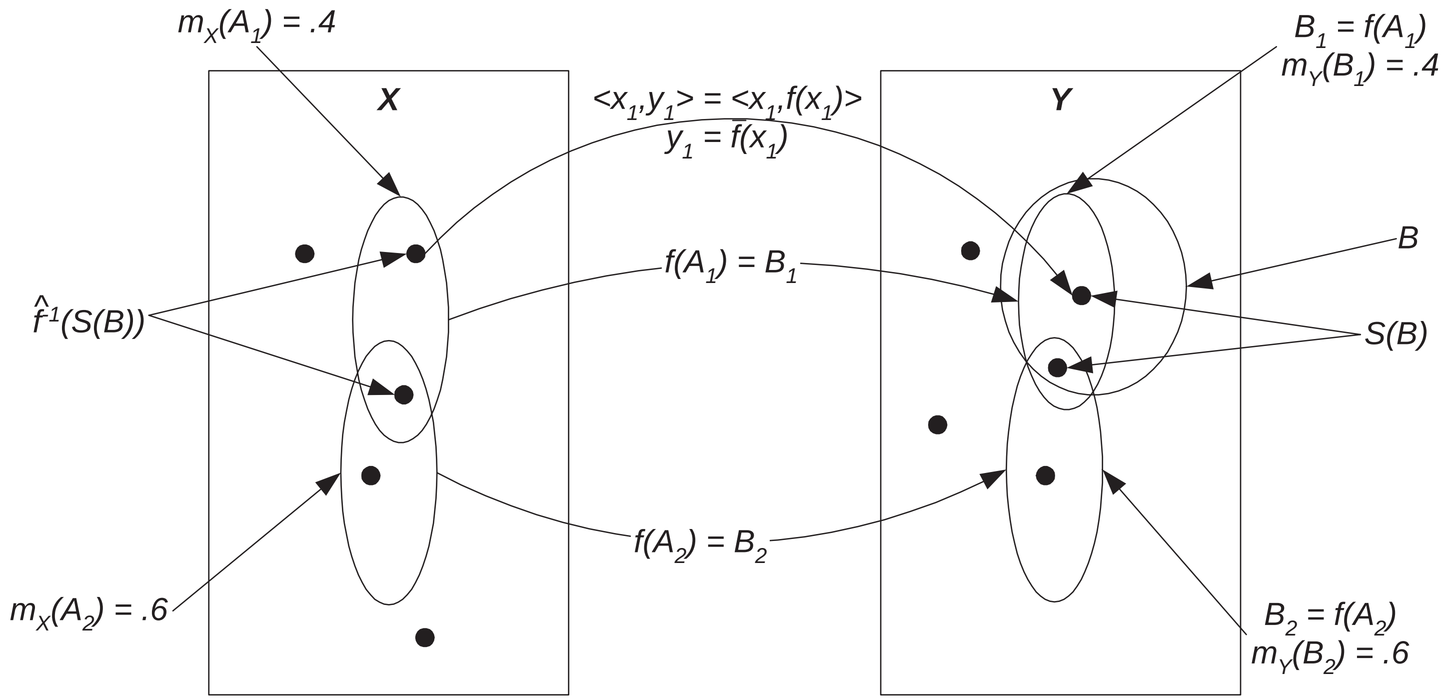


CURRENT PROBLEM

- Set of parameters $X = \{X_i\}$ with
  $$\mathbf{X} := \mathop{\times}_i X_i$$
- Unknown function $f\colon \mathbf{X} \mapsto \mathbf{Y}, f(\vec{x}) = \vec{y}$
- Knowledge on $\mathbf{X}$ characterized by a random interval structure: statistical collection of intervals $\mathcal{A}_i$ on each X_i
- Induces global n-dimensional $\mathcal{A}$ on $\mathbf{X}$
- f induces an output random interval $f(\mathcal{A})$ on $\mathbf{Y}$
- What knowledge of $f(\mathcal{A})$ can be gained by samples of $f(\vec{x})$?

SAMPLING ESTIMATION OF RANDOM SET PROPAGATION

- Initially simplify to finite random sets
- Estimate $\text{Pl}_Y(B)$ by $\widehat{\text{Pl}}_Y(B) := \text{Pl}_X(\hat{f}^{-1}(S(B)))$



BREATHE!

Thank you!