Uncertainty for Qualitative Variables

SAND2007- XXXX

Presented to Consequence Management Team

November 12, 2007

John Darby

6462

Sandia National Laboratories

jldarby@sandia.gov

Sandia National Laboratories

Acknowledgments

- Approximate Reasoning based on Logic Evolved Decision Techniques developed at LANL by Terry Bott and Steve Eisenhawer
- Assistance with Mathematics from Jon Helton, Arizona State/SNL
- Belief/Plausibility for Fuzzy Sets for Numeric Variables: Ron Yager paper
- Adversary Defender Model Concepts
  - Peter Merkle, SNL
  - Brian Bush, LANL
- Graphical Ranking Technique
  - John Cummings, SNL

jldarby@sandia.gov

Sandia National Laboratories

References

- "Evaluation of Risk from Acts of Terrorism: The Adversary/Defender Model using Belief and Fuzzy Sets", SAND2006-5777, September, 2006
- "Evaluation of Risk for Acts of Terrorism using Belief and Fuzzy Sets", Journal of Nuclear Materials Management, Winter, 2007, Volume XXXV, Number 2
- "LinguisticBelief: A Java Application for Linguistic Evaluation using Belief, Fuzzy Sets, and Approximate Reasoning", SAND2007-1299, March, 2007
- "Linguistic Evaluation of Terrorist Scenarios: Example Application", SAND2007-1301, March, 2007

jldarby@sandia.gov

Sandia National Laboratories

Presentation Topics

- **Methodology**
- **Software Tools**

jldarby@sandia.gov

Sandia National Laboratories

Methodology

- Uncertainty
  - Ambiguity
    - Belief/Plausibility Measure
      - Superset of Probability Measure
  - Vagueness
    - Fuzzy Sets
      - Typically used for Numeric Variables
- Qualitative Variables
  - Fuzzy Sets
    - Purely Linguistic Fuzzy Sets
- Combination of Qualitative Variables
  - Convolution of Evidence over Linguistic Fuzzy Sets
    - Approximate Reasoning
    - Belief/Plausibility Measure for Uncertainty

jldarby@sandia.gov

Sandia National Laboratories

Uncertainty

- **Ambiguity**
  - Uncertainty as to what will occur in the future
    - Dow Jones Industrial Average Close on Dec. 31, 2007
      - Will be one value
      - Ambiguity as to what that value will be
- **Vagueness**
  - Uncertainty as how to categorize a known outcome
    - Dow Jones close is 13,876 on Dec. 31, 2007
      - Is this "High" ?
      - What do you mean by "High"?
  - Vagueness can be expressed Linguistically (Words)

jldarby@sandia.gov

Sandia National Laboratories

Ambiguity: Aleatory and Epistemic

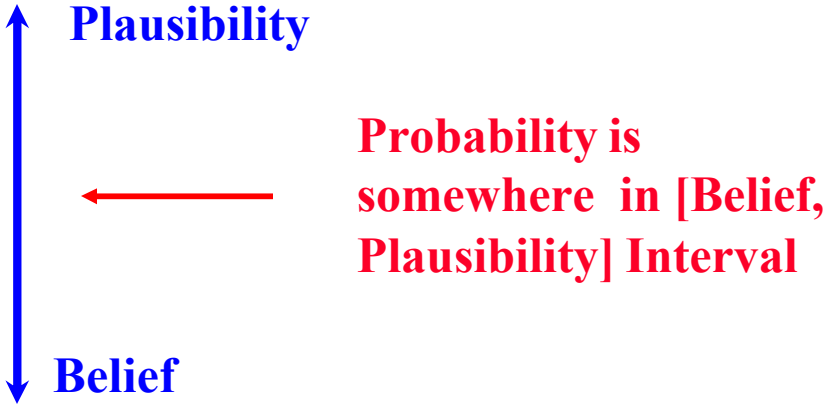
- For a Fair Coin
  - Uncertainty is Aleatory (random)
  - Probability Heads is ½
  - Probability Tails is ½
- But if we cannot toss coin, we do not know coin is fair, we do not even knows if coin has Heads and Tails
  - May not be Fair Coin (may be Weighted for Tails)
  - May be Two-Headed or Two-Tailed Coin
  - Epistemic (state of knowledge) uncertainty
  - Insufficient information to assign Probability to Heads and Tails
  - For Total Ignorance
    - Belief/Plausibility for Heads is 0/1
    - Belief/Plausibility for Tails is 0/1
- With more information (actually tossing the coin) we can reduce Epistemic Uncertainty
  - If at least one Heads and one Tails occur in a series of tosses, we know coin has Heads and Tails
  - Many tosses needed to assess if coin is fair
- For Fair Coin we cannot reduce aleatory uncertainty

jldarby@sandia.gov

Sandia National Laboratories

Belief and Plausibility

- Belief / Plausibility form a Lower / Upper Bound for Probability
- Belief is what probability will be
- Plausibility is what probability could be



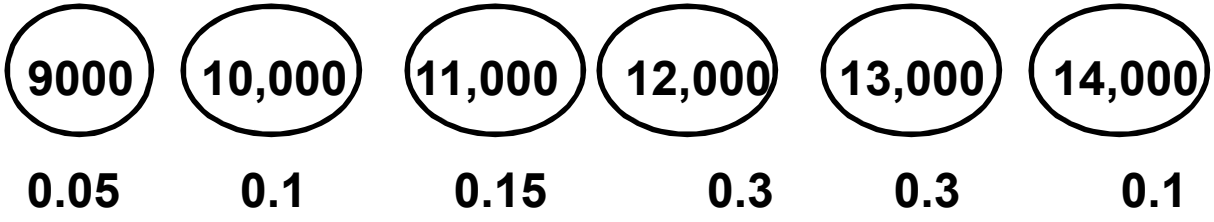
- Similar to a Confidence Interval for a Parameter of a probability distribution; a confidence measure that parameter is in interval, but exactly where in interval is not known
- Belief/Plausibility both reduce to Probability if Evidence is Specific

jldarby@sandia.gov

Sandia National Laboratories

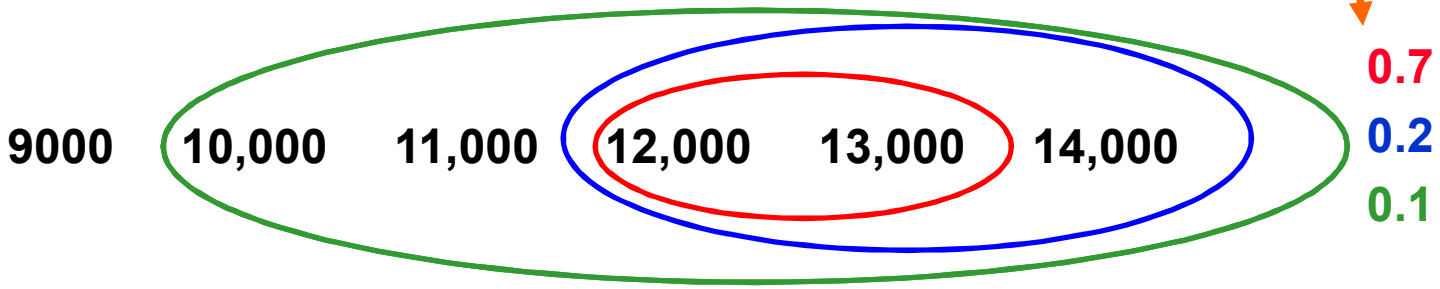
Example of Evidence: Predict Stock Market Close Dec. 31, 2007

- **Probability**



Evidence

- **Belief/Plausibility**



jldarby@sandia.gov

Sandia National Laboratories

Why is this useful?

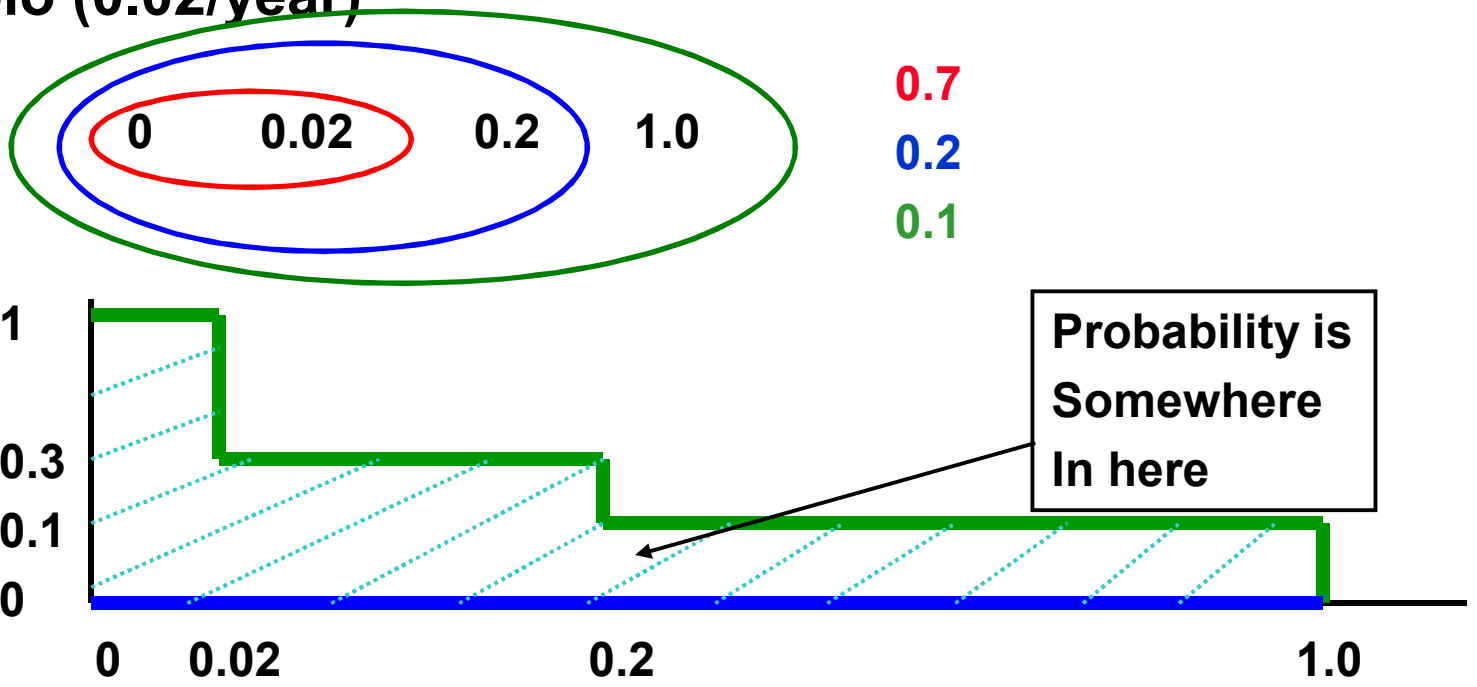
- What is Likelihood of Bio-Terror attack against a Major US City?
  - Frequency of Attack (per year) somewhere in [0, 1]
  - Point Estimate? Useless, too much uncertainty
  - Probability Distribution?
    - Must assign probability to each value in [0, 1]
      - Probability for 0 attacks per year is high
      - Probability assigned to 0 affects probabilities for all other values as must sum to 1.0
        - Is P(0) = 0.01, 0.001, 0.00001, …? Don't Know

jldarby@sandia.gov

Sandia National Laboratories

Why is this useful?

- Evidence is about 1 major Attack every 5 years (0.2/year)
  - Assume Expert Opinion is: 10% Chance Attack is bio (0.02/year)



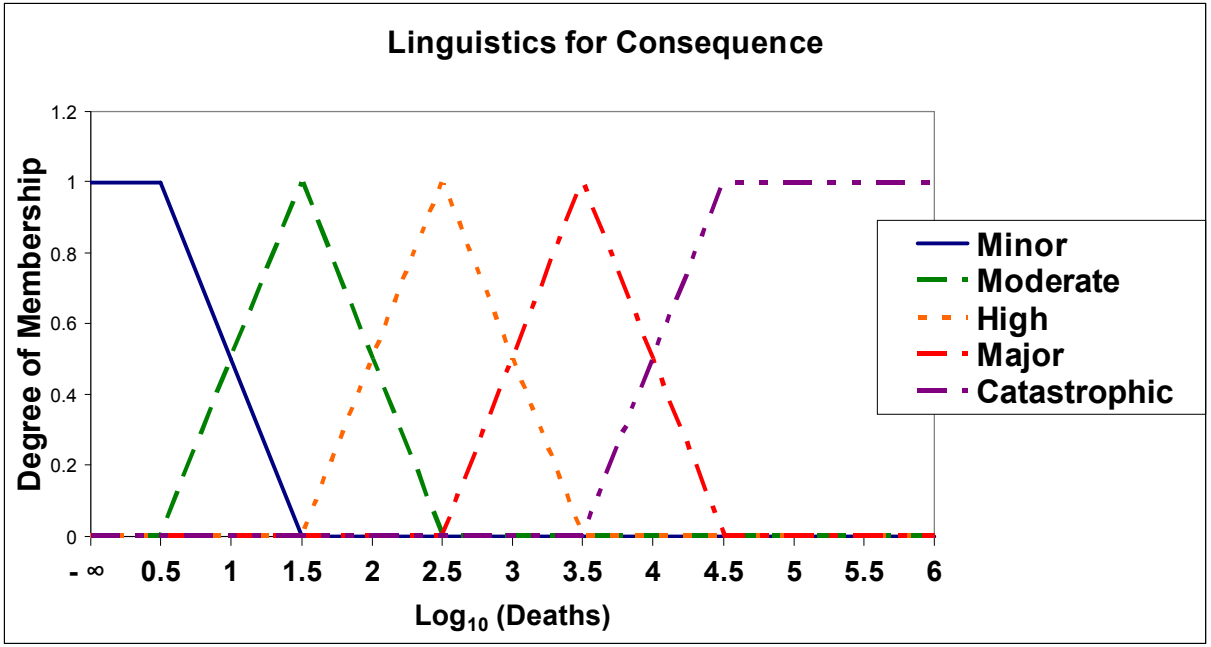
Automated in BeliefConvolution Java code

jldarby@sandia.gov

Sandia National Laboratories

Vagueness
Fuzzy Sets for Numeric Variable

- Represent Variable with Sets to reason at Fidelity Desired. Above 30,000 deaths is "Catastrophic".
- Use Fuzzy Sets to Avoid Sharp Distinction. "Major" Deaths is Between *About* 1000 and *About* 10,000. 999 and 1001 deaths are each part "High" and part "Major".

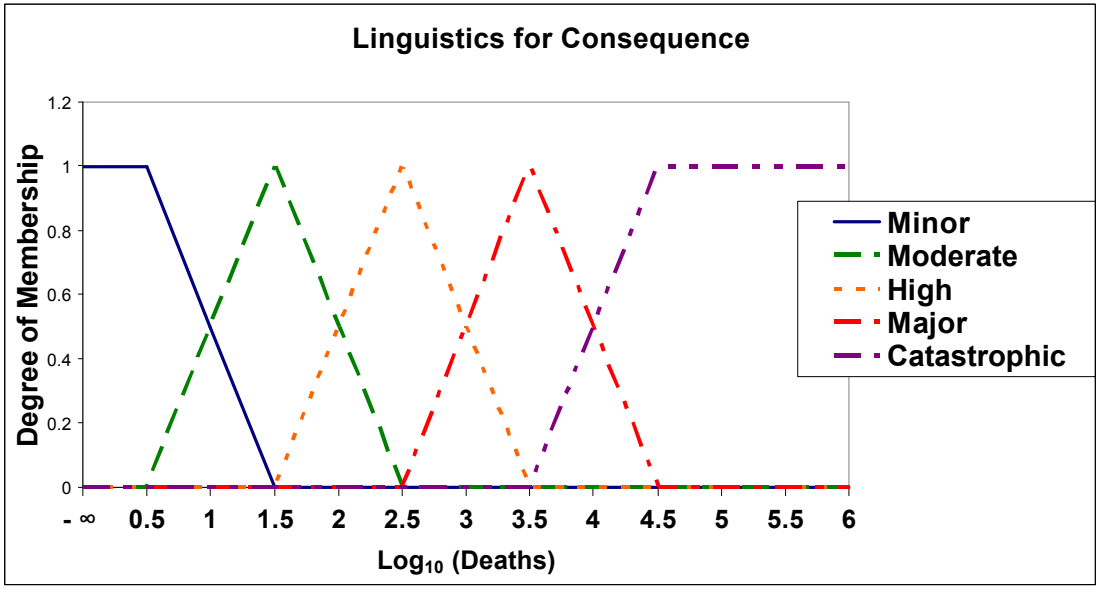


jldarby@sandia.gov

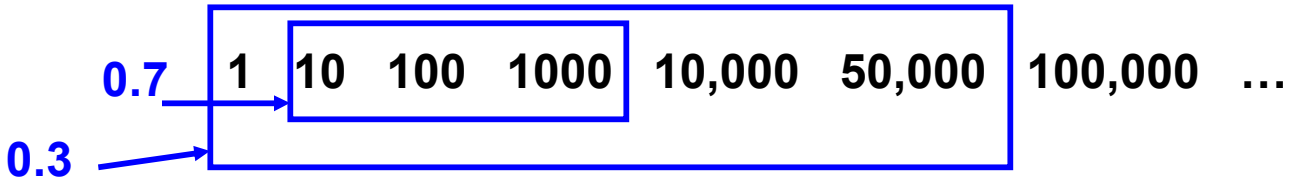
Sandia National Laboratories

Uncertainty for Fuzzy Sets: Numeric Variable

Fuzzy Sets for Deaths



Evidence For Deaths



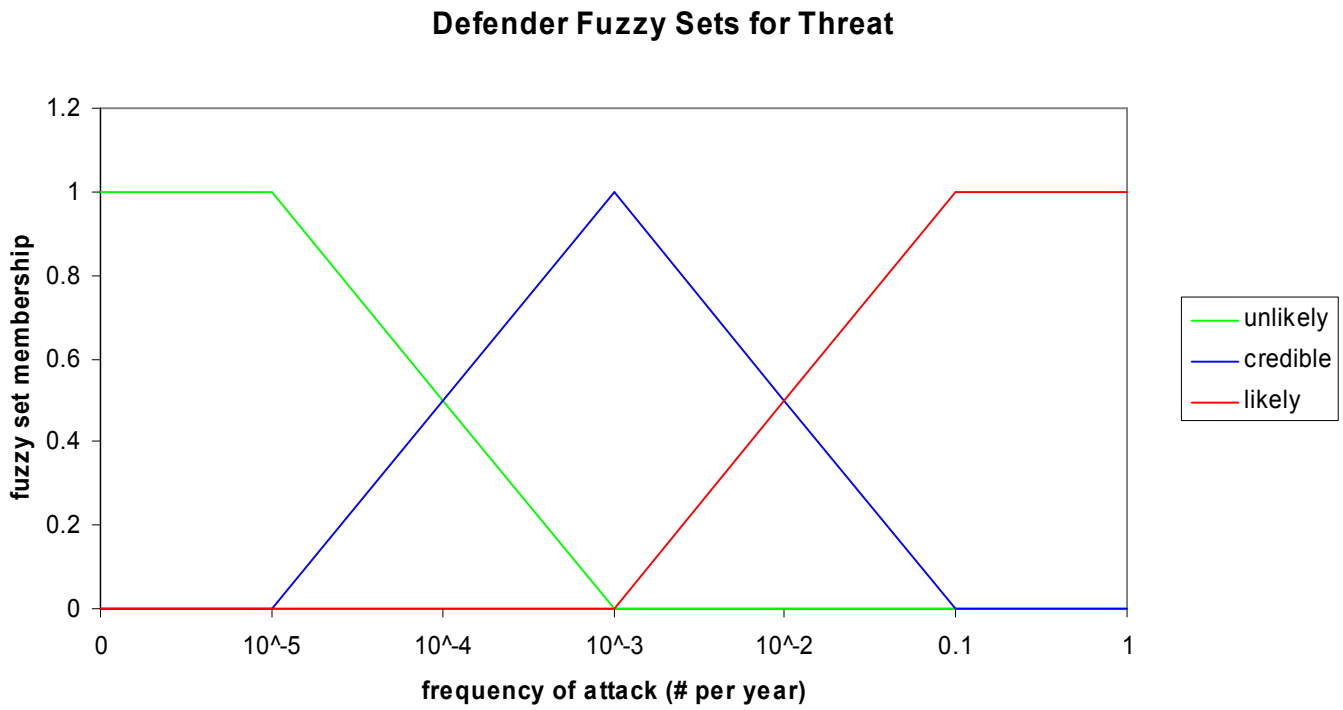
Uncertainty Distribution for Deaths:

	Minor	Moderate	High	Major	Catastrophic
Belief / Plausibility	0 / 0.65	0 / 1	0 / 1	0 / 0.65	0 / 0.3

(Calculated with BeliefConvolution code using Yager Method)

Why is this Important?

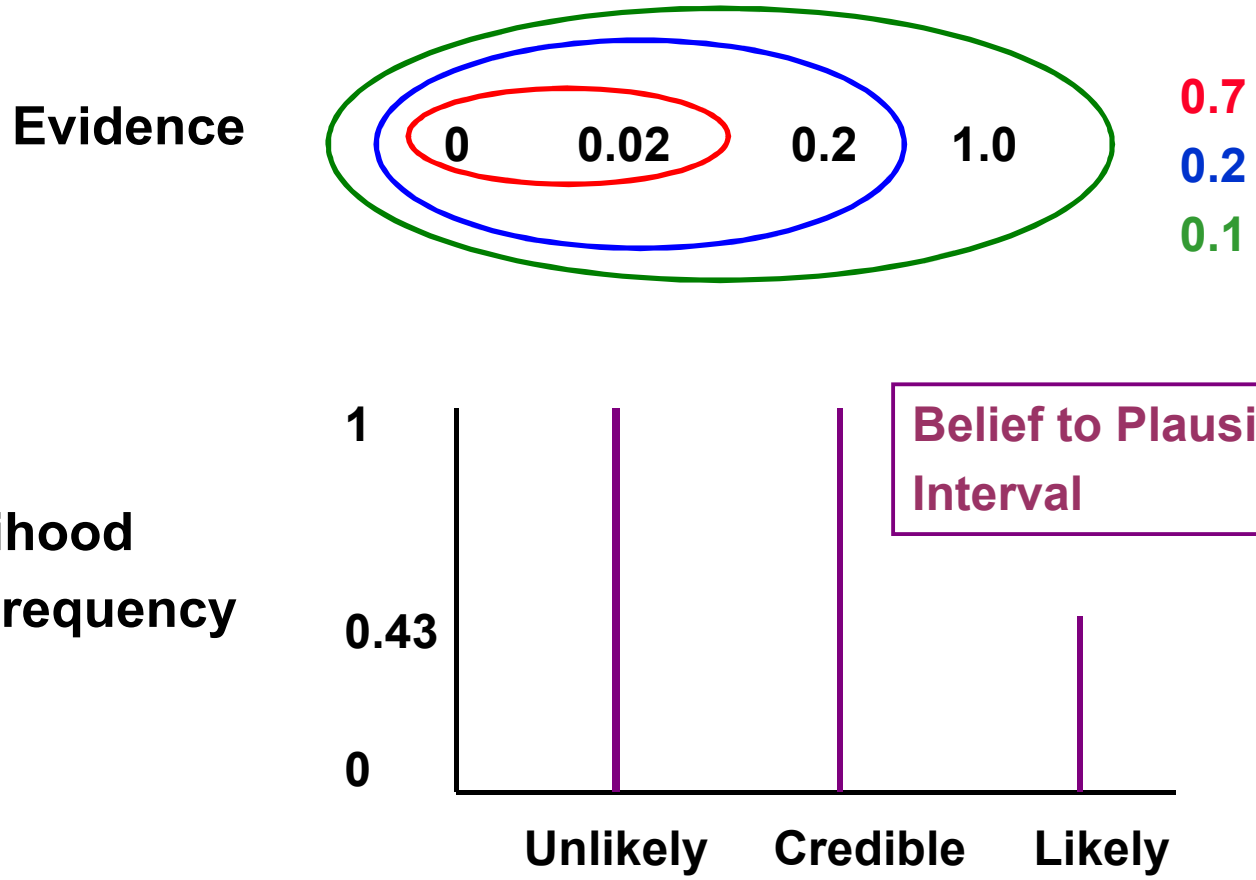
- What is Likelihood of Bio-Terror attack against a Major US City?
  - Evidence is about 1 major Attack every 5 years (0.2/year)
    - Assume Expert Opinion is: 10% Chance Attack is bio (0.02/year)
- Assume Following Fuzzy Sets for Evaluating Frequency of Attack



jldarby@sandia.gov

Sandia National Laboratories

Why is this Important?



(Calculated with BeliefConvolution code using Yager Method)

jldarby@sandia.gov

Sandia National Laboratories

Qualitative Variables

- Variable Segregated into Purely Linguistic Fuzzy Sets
  - Variable: "Health"
  - Fuzzy Sets: "Bad", "Moderate", "Excellent"
- Why Pure Linguistics?
  - Numeric Scale is Unknown
    - Is "Health" [0, 10], [0, 10^6], [-700, square root of 42]?
  - Scaling is Un-Manageable when Combine Variables
    - Combine "Health" with "Wealth" to Evaluate "Quality of Life"
      - "Wealth" can be Numeric: [$0, $50B]
      - What is Numeric Scale for "Health"?
      - What is Numeric Scale for "Quality of Life"?

jldarby@sandia.gov

Sandia National Laboratories

Fuzzy Sets for Non-Numeric Variable

Adversary Level of Technical Training:

- **High School**
- **Bachelors**
- **Advanced**

Do NOT Force Numeric Measure: Requires Arbitrary Scale

Adversary Level of Technical Training:

- **High School = 1?**
- **Bachelors = 2?**
- **Advanced = 3?**

Adversary Level of Technical Training:

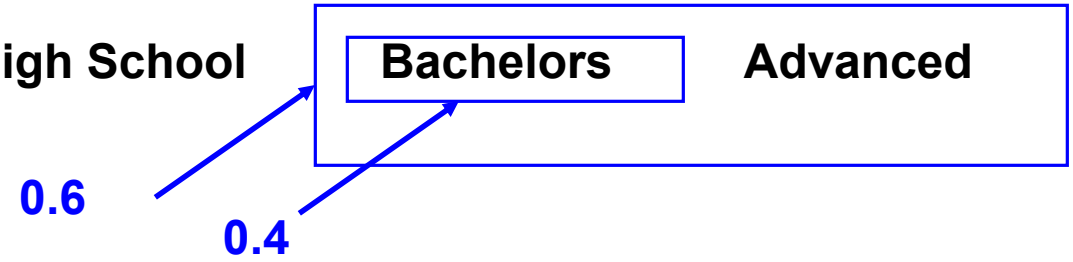
- **High School = 10?**
- **Bachelors = 100?**
- **Advanced = 1000?**

jldarby@sandia.gov

Sandia National Laboratories

Uncertainty for Fuzzy Sets: Non-Numeric Variable

- **Fuzzy Sets for Adversary Level of Technical Training**
  - **High School Bachelors Advanced**
- **Evidence**



- **Uncertainty Distribution: Belief / Plausibility**

High School	Bachelors	Advanced
0/ 0	0.4 / 1	0 / 0.6

jldarby@sandia.gov

Sandia National Laboratories

Combining Variables: Convolution of Uncertainty Distributions

- **Belief/Plausibility Distributions**
  - Evidence Over Fuzzy Sets for Each Variable
- **Convolute Distributions** per the Rule Base
  - Mathematics of Belief/Plausibility
- Same Concept as Convolution of Probability Distributions
  - Mathematics of Probability

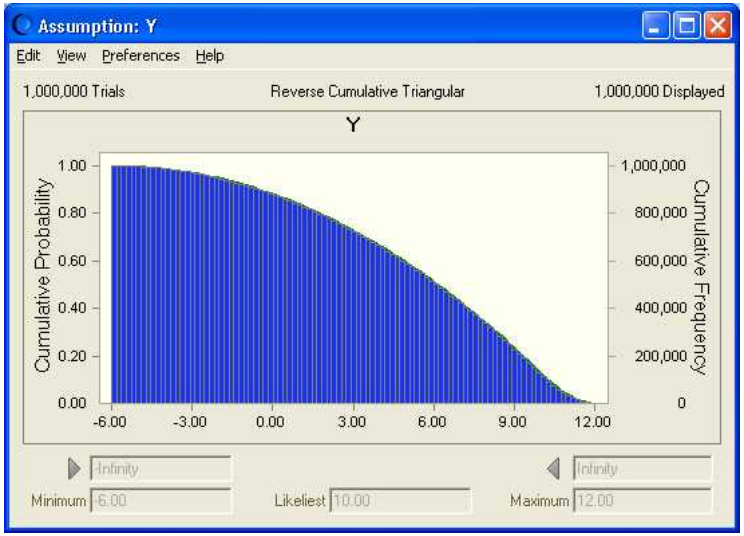
jldarby@sandia.gov

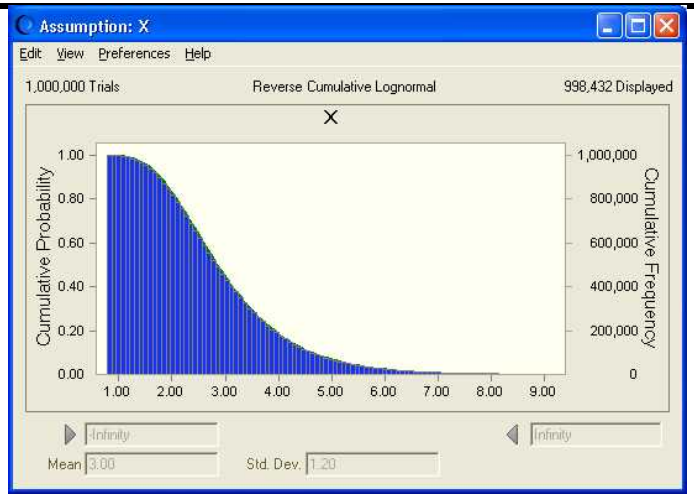
Sandia National Laboratories

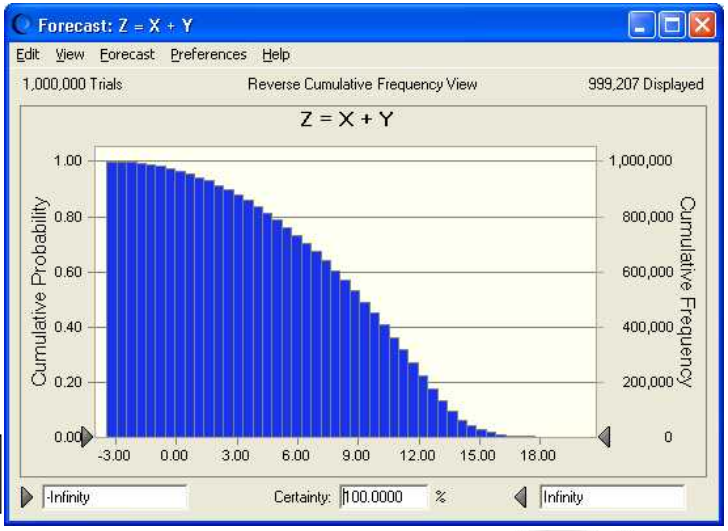
Convolute Probability Distributions: Crystal Ball Software

Z = X + Y

X and Y Independent







Mean of Z is a Point value: 8.34

jldarby@sandia.gov

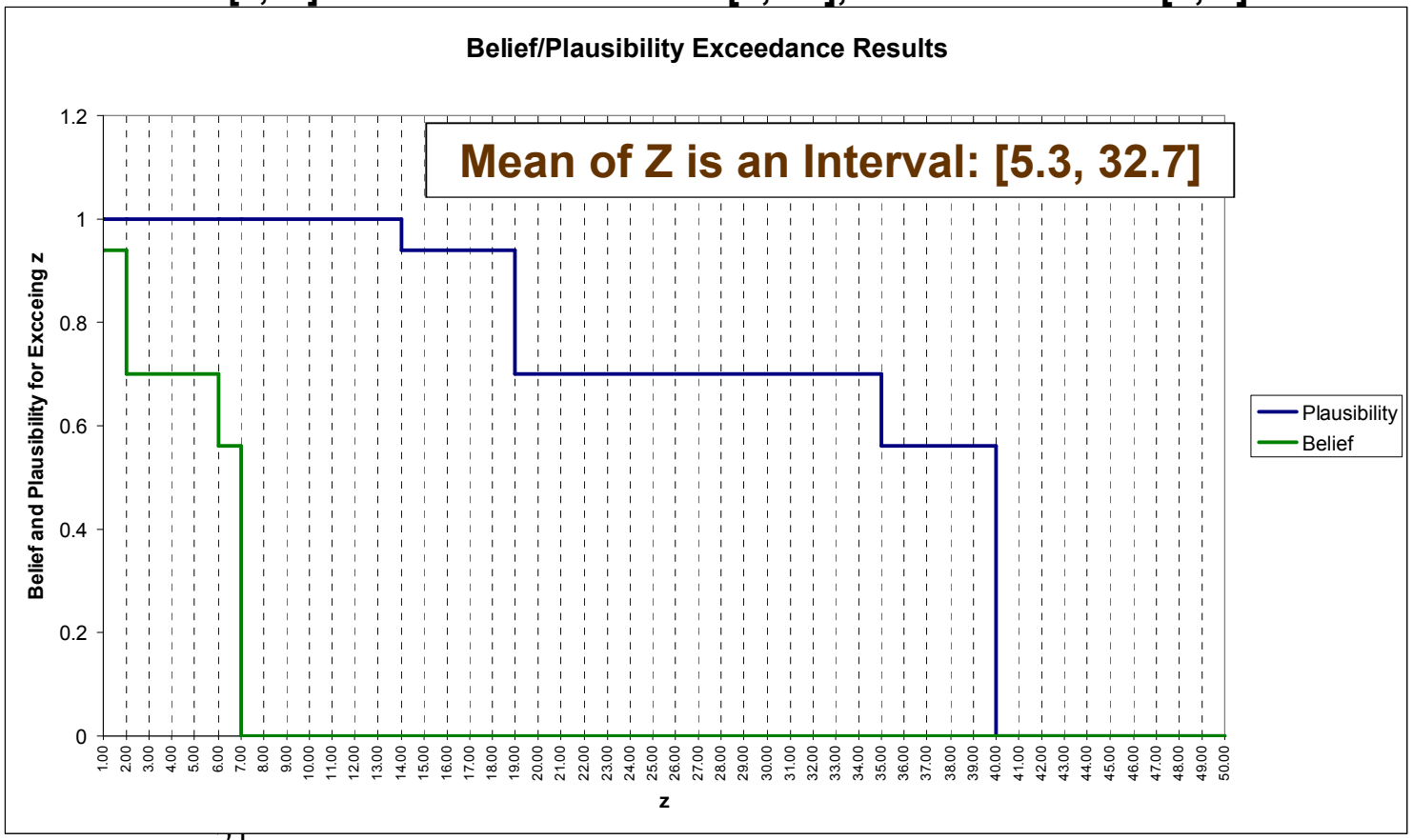
Sandia National Laboratories

Convolute Belief/Plausibility Distributions for Numeric Variables: BeliefConvolution Software

Z = X + Y, X and Y Non-Interactive

X over [1, 20] with Evidence: 0.8 for [2, 15], 0.2 for [1,10]

Y over [0,30] with Evidence: 0.7 for [5, 25], Evidence 0.3 for [0, 4]



jldarby@sandia.gov

Sandia National Laboratories

Convolute Belief/Plausibility Distributions for Linguistic Variables: LinguisticBelief Software

- **Example Follows**

jldarby@sandia.gov

Sandia National Laboratories

Combining Qualitative Variables: Approximate Reasoning

- Mathematics for Combining Words
- If we use Words instead of numbers we need a way of combining the Words for Different Variables
- Implemented as A Rule Base for Combining Fuzzy Sets from Different Variables

jldarby@sandia.gov

Sandia National Laboratories

Combination of Linguistic Variables: Example

Develop the Model: Happiness for Any Individual

Define the Variables and their Fuzzy Sets

- **Basic Variables**
  - Health
    - Bad, Moderate, Excellent
  - Wealth
    - Poor, Middle Class, Rich
  - Outlook on Life
    - Pessimist, Optimist
- **Rule Based Variables**
  - Quality of Life = Heath x Wealth (x per rule base)
    - Not so Good, Good
  - Happiness = Outlook on Life x Quality of Life
    - Depressed, Accepting, Very Happy

jldarby@sandia.gov

Sandia National Laboratories

Combination of Linguistic Variables: Example

Develop the Approximate Reasoning Rule Base for Rule Based Variables

Quality of Life

Rules for selected RuleLinguistic

Rules for RuleLinguistic: Quality of Life

Fuzzy Set for Input Linguistic: Health	Fuzzy Set for Input Linguistic: Wealth	Output Fuzzy Set for Rule (blank if rule not set)
Bad	Poor	Not so Good
Bad	Middle Class	Not so Good
Bad	Rich	Not so Good
Moderate	Poor	Not so Good
Moderate	Middle Class	Not so Good
Moderate	Rich	Good
Excellent	Poor	Good
Excellent	Middle Class	Good
Excellent	Rich	Good

Specify Output Fuzzy Set for Selected Rule Choices Are:

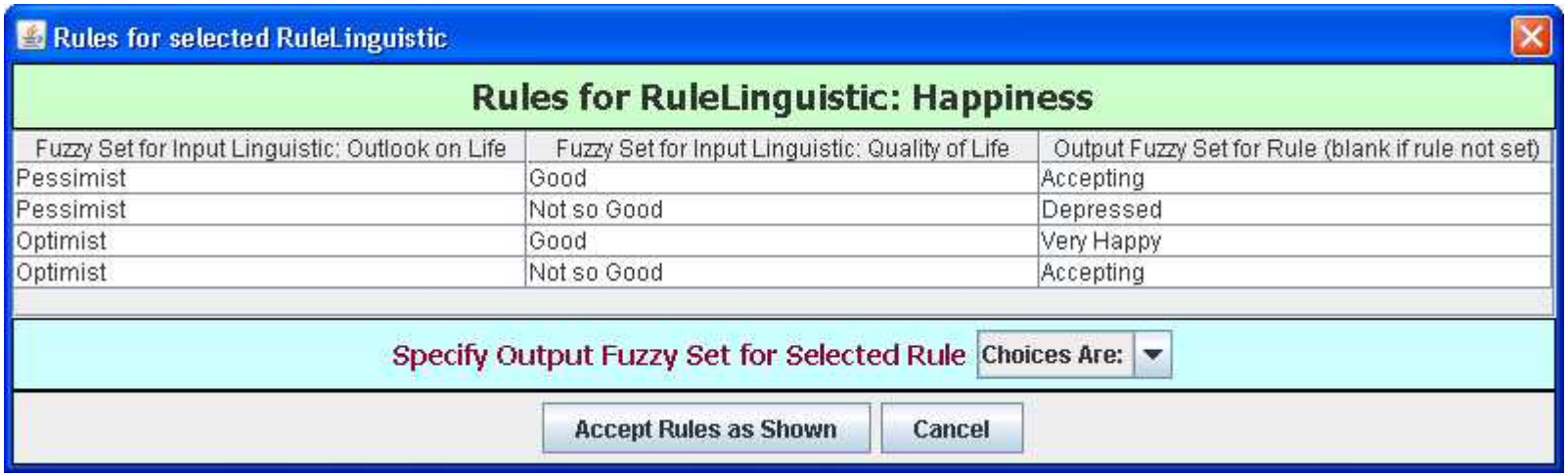
Accept Rules as Shown Cancel

jldarby@sandia.gov

Sandia National Laboratories

Combination of Linguistic Variables: Example

Happiness



jldarby@sandia.gov

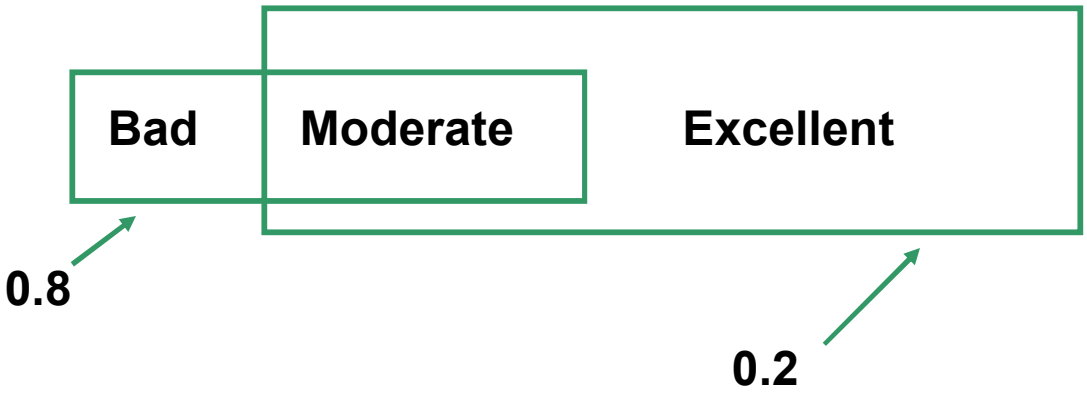
Sandia National Laboratories

Combination of Linguistic Variables: Example

Evaluate the Model for Specific Individual: Happiness for "John"

Assign Evidence to Fuzzy Sets for Basic Variables

Health

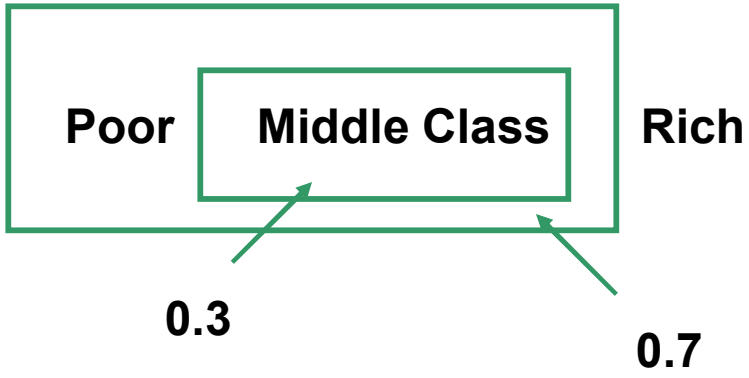


jldarby@sandia.gov

Sandia National Laboratories

Combination of Linguistic Variables: Example

Wealth



jldarby@sandia.gov

Sandia National Laboratories

Combination of Linguistic Variables: Example

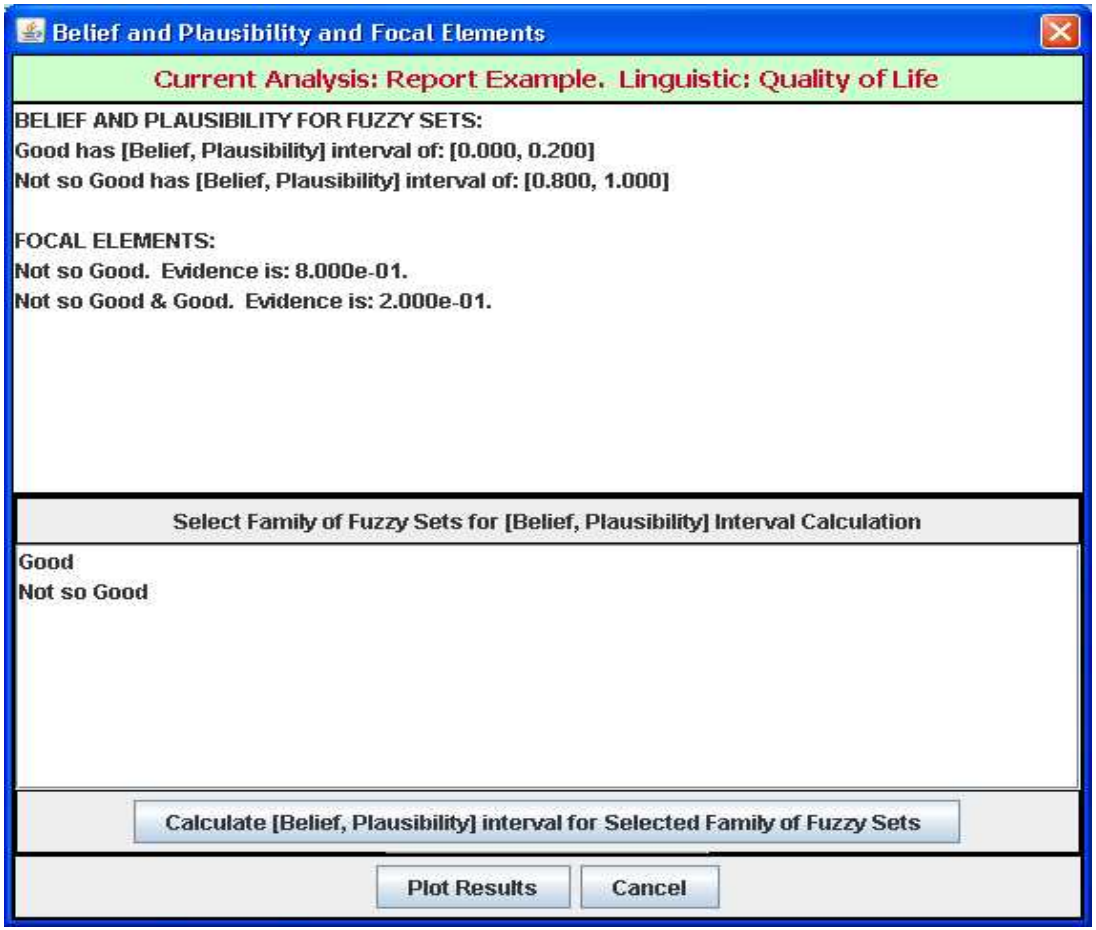
Outlook on Life

jldarby@sandia.gov

Sandia National Laboratories

Combination of Linguistic Variables: Example

Evaluate Variable: Quality of Life for John

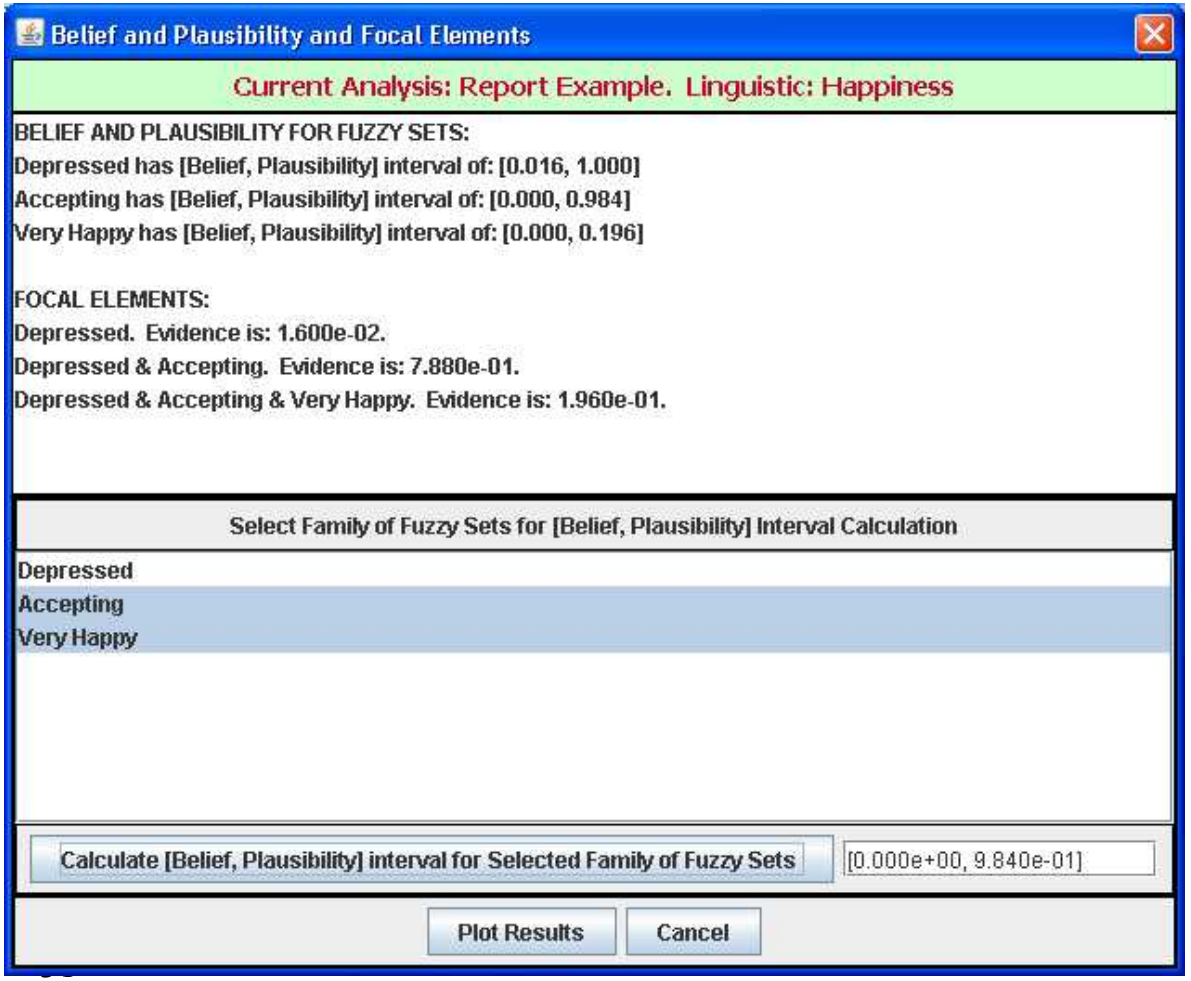


jldarby@sandia.gov

Sandia National Laboratories

Combination of Linguistic Variables: Example

Evaluate Variable Happiness for John

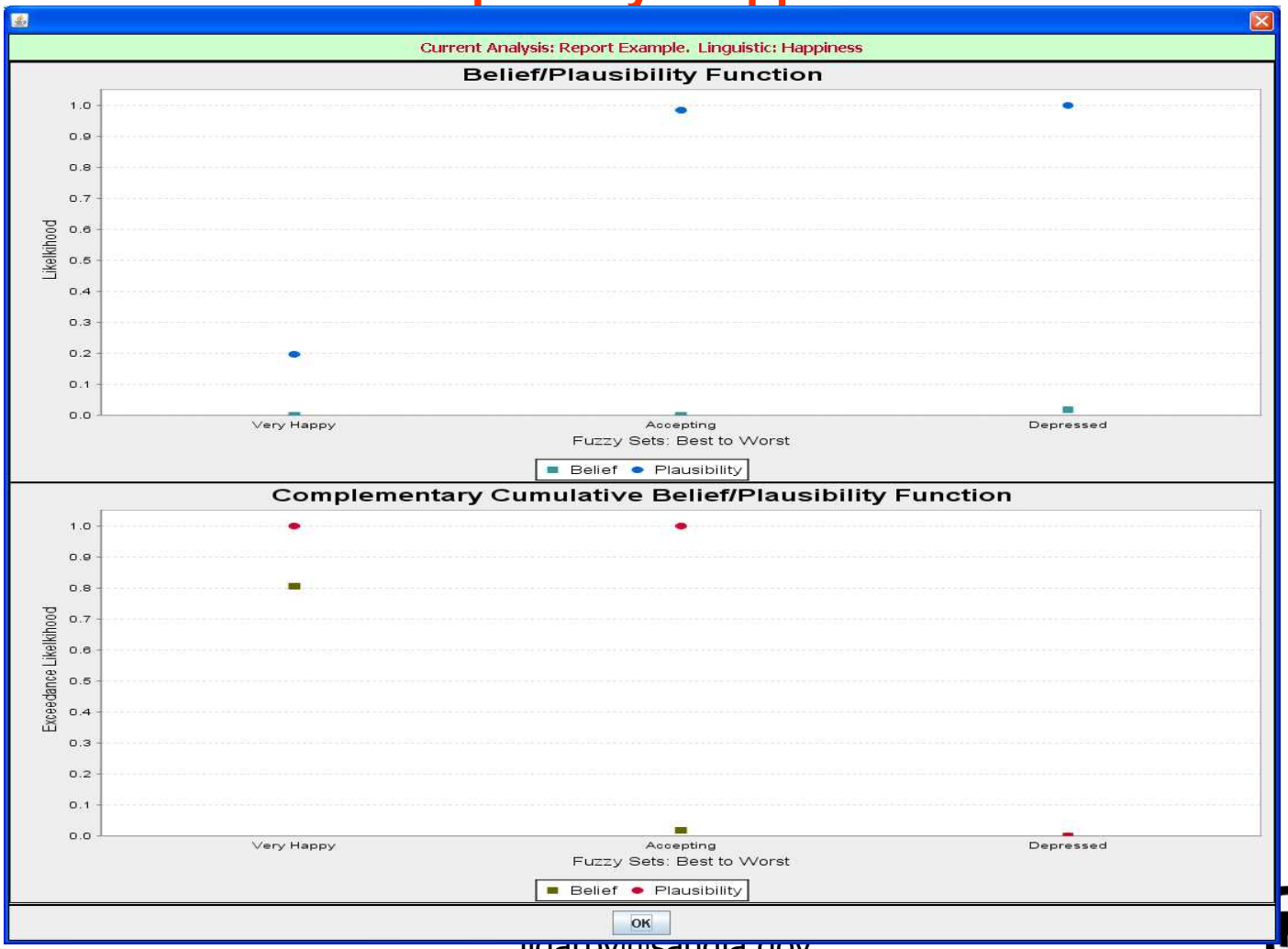


jldarby@sandia.gov

Sandia National Laboratories

Combination of Linguistic Variables: Example

Summarize Results Graphically: Happiness for "John"



jdarby@sandia.gov

Sandia National Laboratories

Results Expressed as Complementary Cumulative Belief/Plausibility Functions (CCBPFs) for Linguistic Variable

- Linguistic Fuzzy Sets Ordered from "Best" to "Worst"
  - CCBPFs are Non-Increasing
- "Likelihood" of Exceeding Fuzzy Set
- "Likelihood" is Belief/Plausibility Interval
- Analogous to Complementary Cumulative Probability Function (CCPF) for Random Variable
  - CCPF Random Variable is a real number
    - discrete or continuous
  - CCBPFs Variable has linguistic fuzzy sets
    - Discrete
  - CCPF is a One Function: a Curve
  - CCBPFs are Two Functions: an Interval

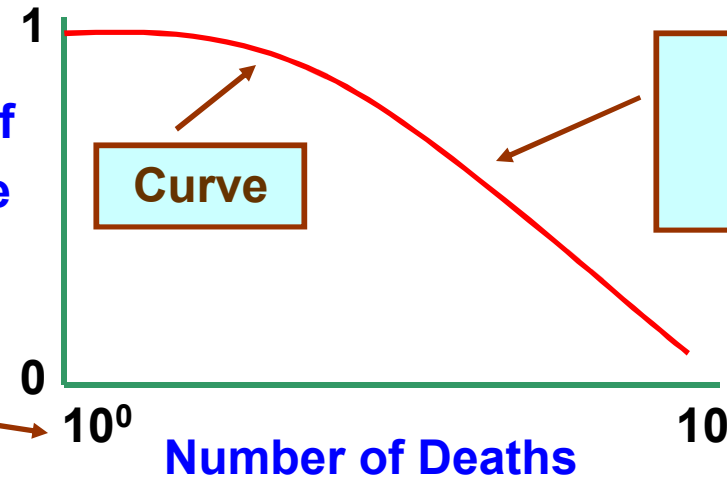
jldarby@sandia.gov

Sandia National Laboratories

CCPF and CCBPFs

- CCPF

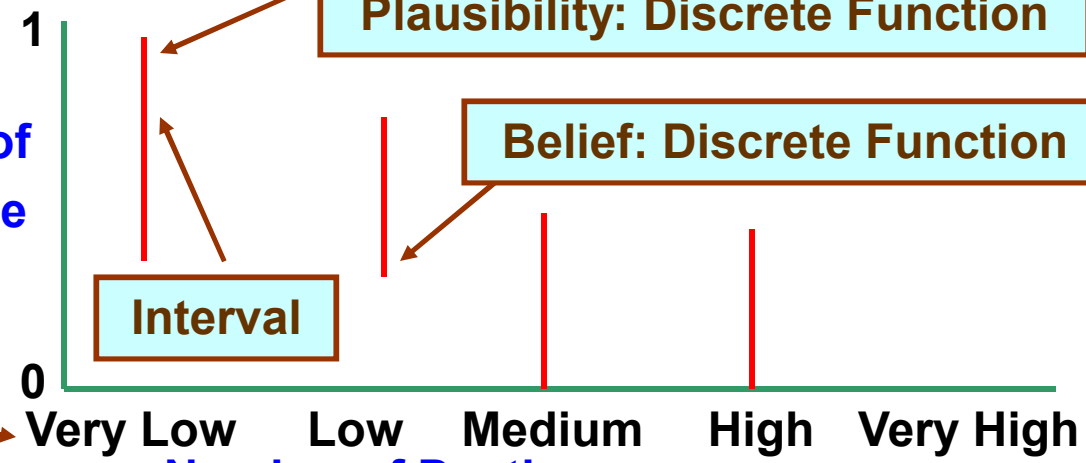
Likelihood of Exceedance



Continuous Variable: Real Number

- CCBPFs

Likelihood of Exceedance



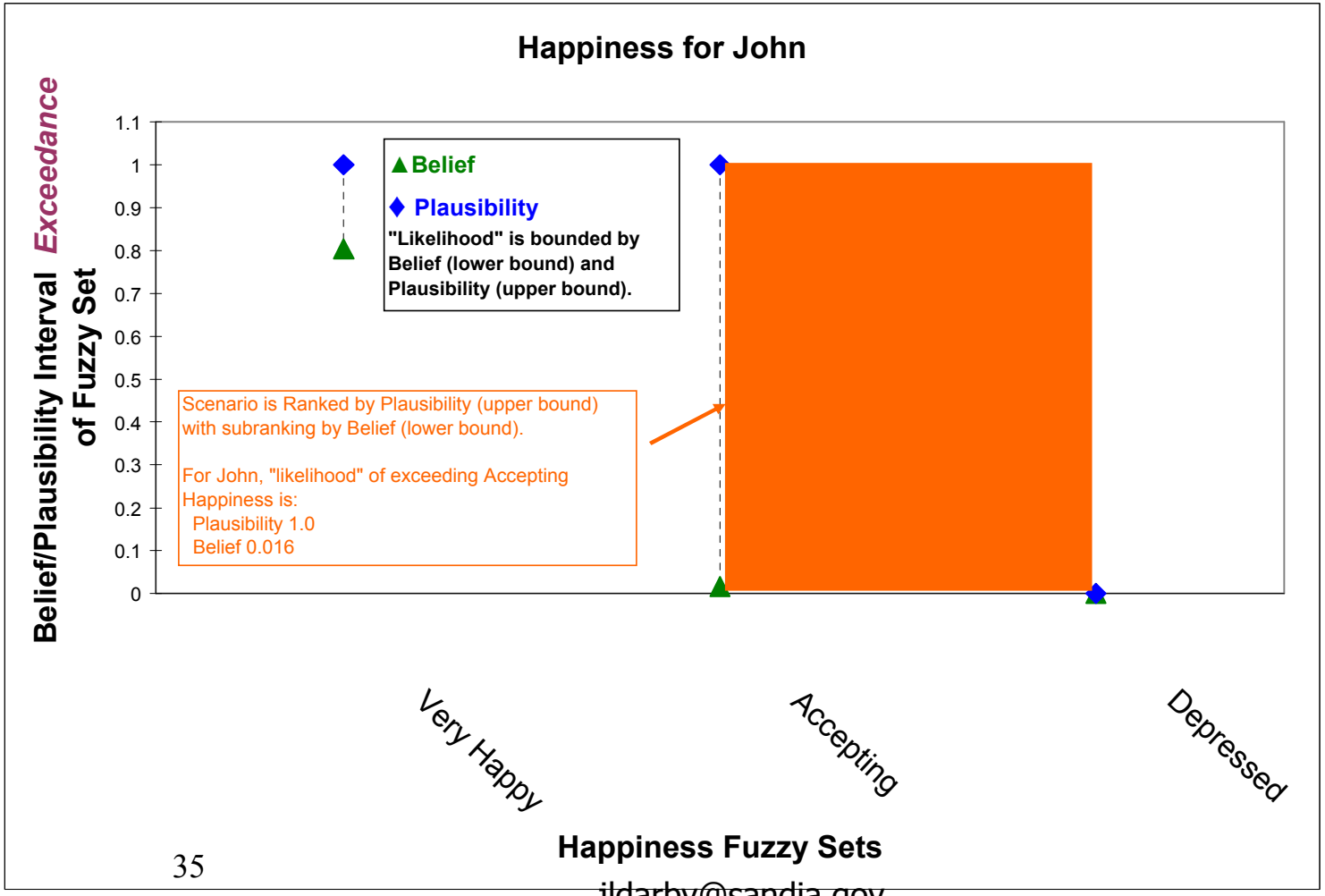
Discrete Variable: Linguistic Fuzzy Sets

jldarby@sandia.gov

Sandia National Laboratories

Combination of Linguistic Variables: Example

Graphical Summary for Ranking: "John"



jldarby@sandia.gov

Sandia National Laboratories

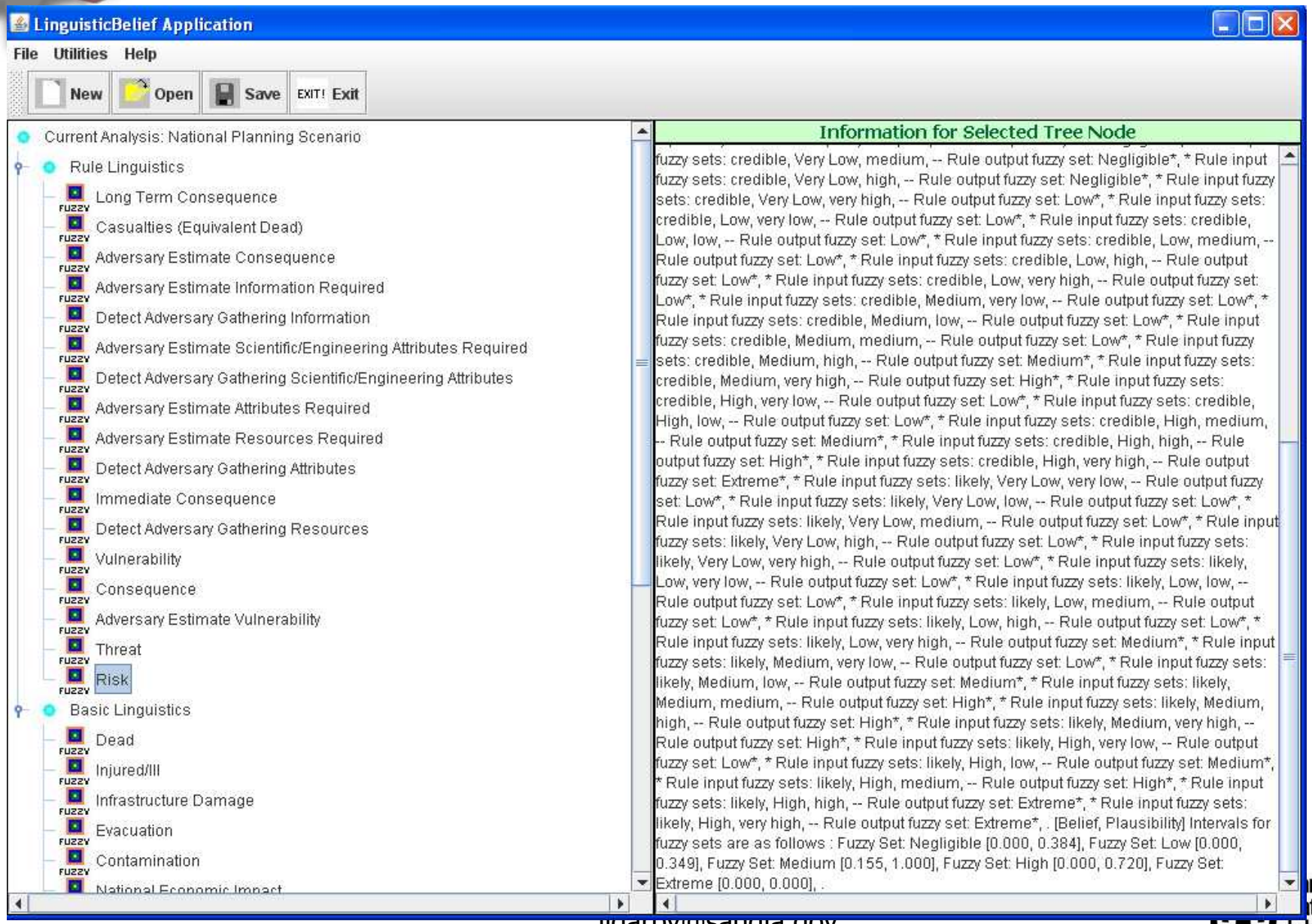
Custom Software Tools: Java

- BeliefConvolution
  - Convolution of Numeric Variables with Belief/Plausibility
- LinguisticBelief
  - Evaluation of Linguistic Variables
    - Linguistic Fuzzy Sets
    - Approximate Reasoning
    - Belief/Plausibility
- PoolEvidence
  - Multiple Experts provide Evidence for Variables
    - Linguistic Fuzzy Sets
  - Combine Evidence
    - Pooled Evidence for Variables
  - Input for LinguisticBelief

jldarby@sandia.gov

Sandia National Laboratories

LinguisticBelief: Example Application



jdarby@sandia.gov

Sandia National Laboratories

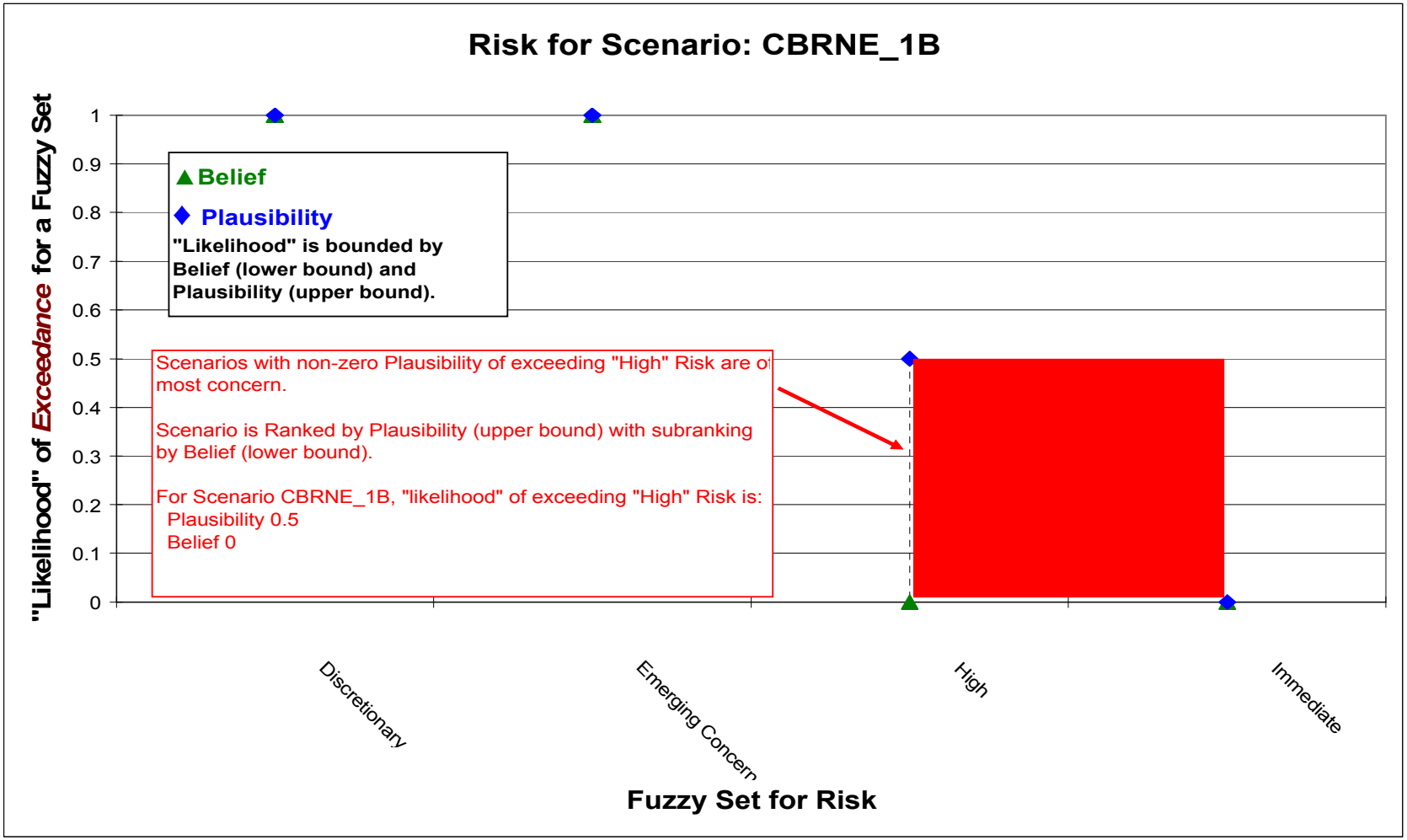
Rank Order Scenarios by Risk

- Rank a scenario by the Highest Non-Zero Plausibility of Exceeding the "Worst" Fuzzy Set
  - For Scenarios with Equal Plausibility, Subrank by Highest Belief
  - Extension of "Probability of Exceedance" approach
    - Uses Fuzzy Sets instead of Numbers
    - Uses Belief/Plausibility Interval instead of Probability
  - Can be "Color Coded"
    - Shown for 3 of 5 scenarios in Following from SAND2007-1301

jldarby@sandia.gov

Sandia National Laboratories

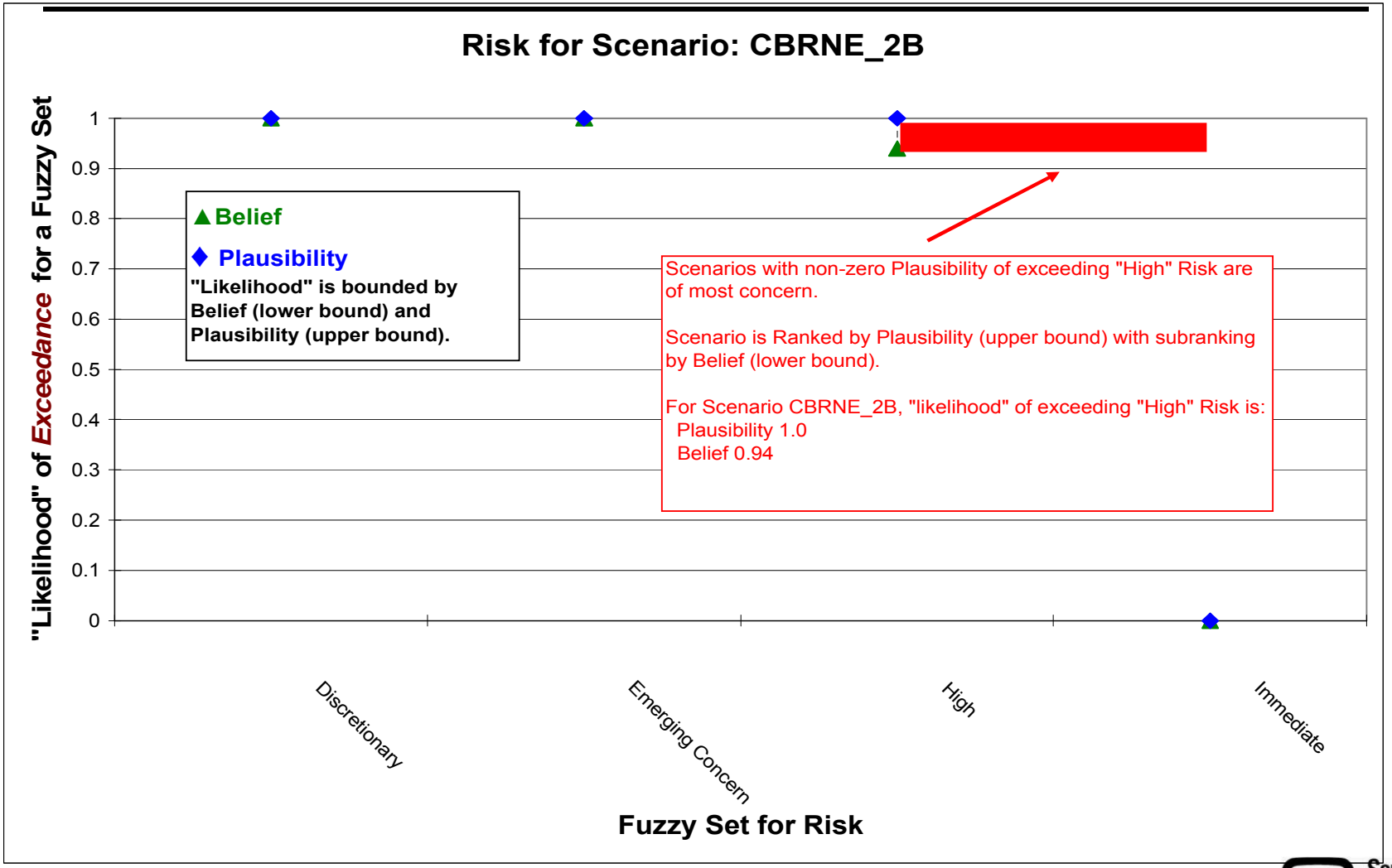
Rank Order Results



jldarby@sandia.gov

Sandia National Laboratories

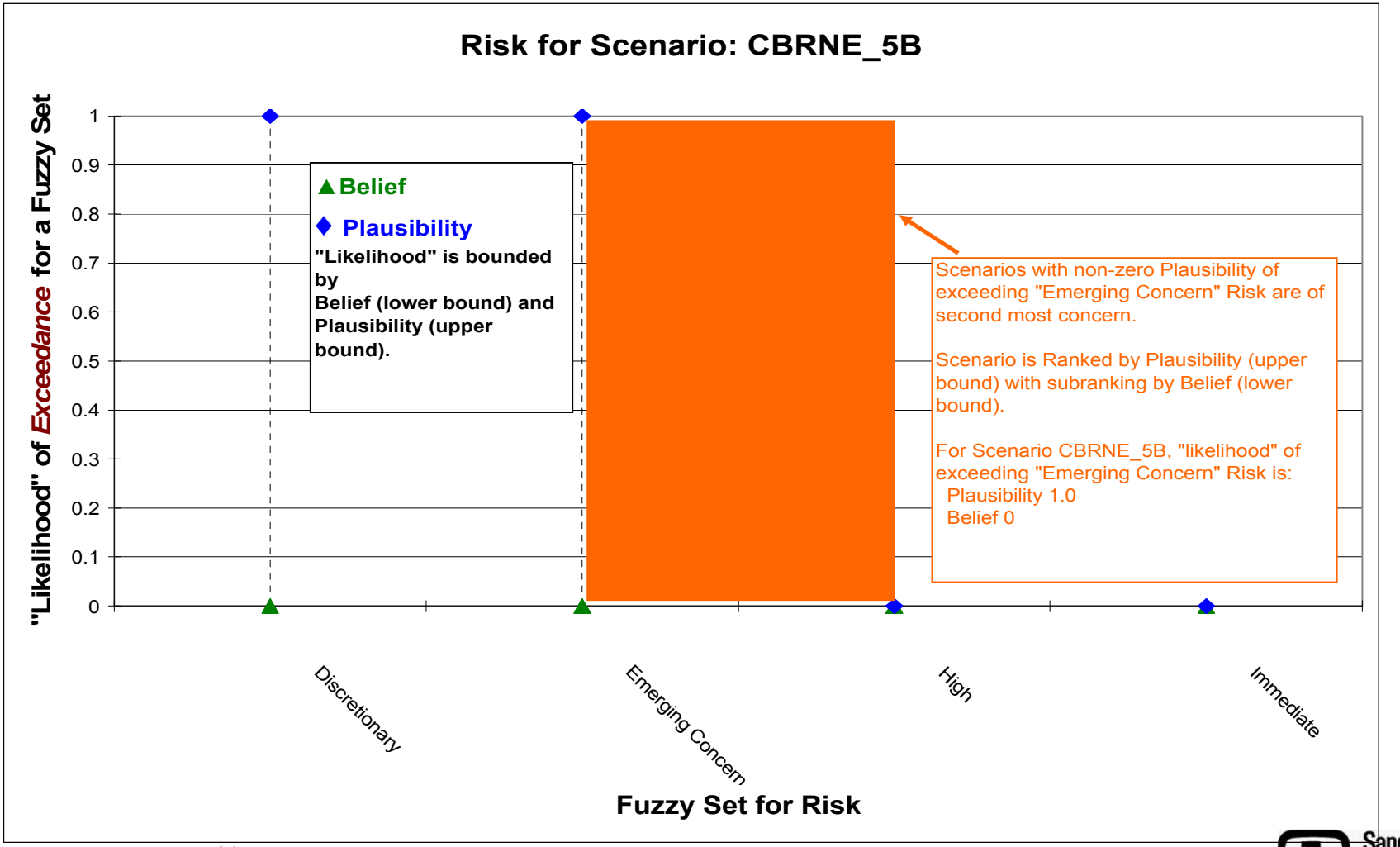
Rank Order Results



jldarby@sandia.gov

Sandia National Laboratories

Rank Order Results



jldarby@sandia.gov

Sandia National Laboratories

Results of Ranking of All Five Scenarios

RANKING FOR SCENARIOS CBRNE_1B through CBRNE_5B

For *Exceeding* Fuzzy Set "High" the Scenarios rank ordered (decreasing) are:
CBRNE_2B has plausibility of exceedance of 1.0 and belief of exceedance of 0.94
CBRNE_3B has plausibility of exceedance of 1.0 and belief of exceedance of 0.77
CBRNE_4B has plausibility of exceedance of 1.0 and belief of exceedance of 0.64
CBRNE_1B has plausibility of exceedance of 0.5 and belief of exceedance of 0.0

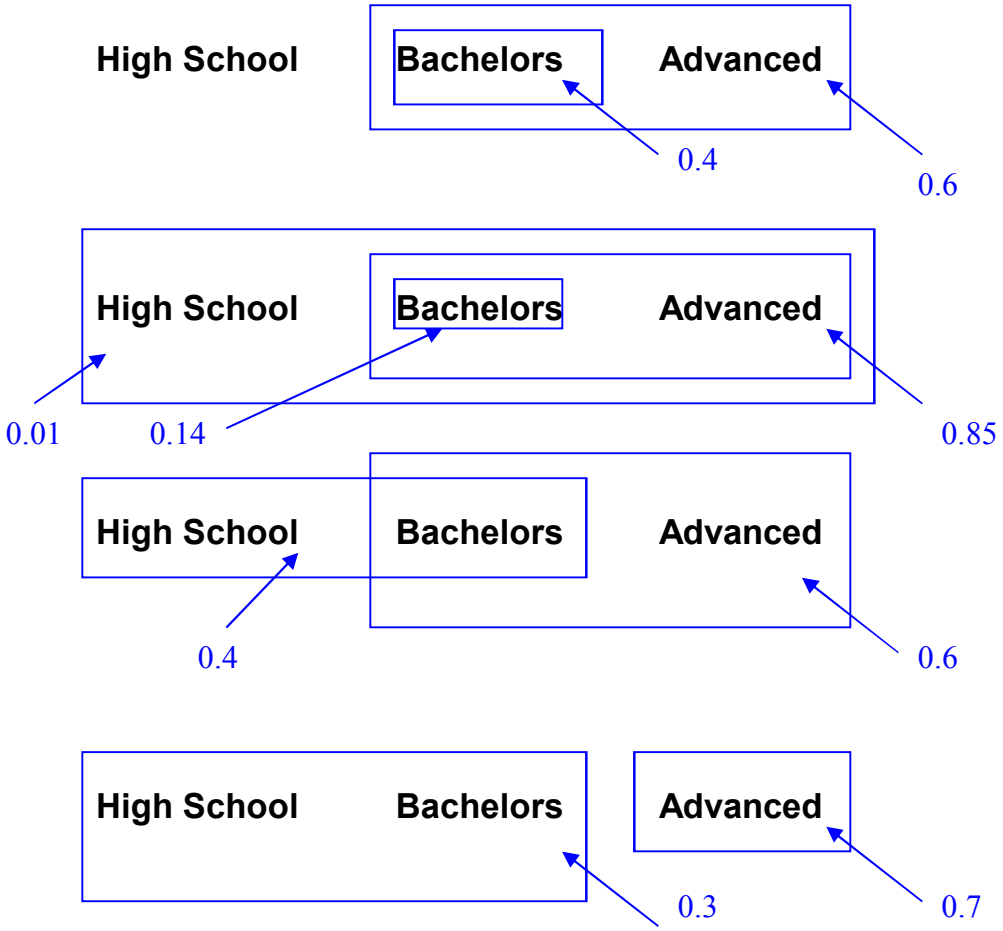
For *Exceeding* Fuzzy Set "Emerging Concern" the Scenarios rank ordered (decreasing) (not already ranked for a worse fuzzy set) are:
CBRNE_5B has plausibility of exceedance of 1.0 and belief of exceedance of 0.0

jldarby@sandia.gov
Sandia National Laboratories

Pool Evidence from Many Experts

Adversary Level of Technical Training:

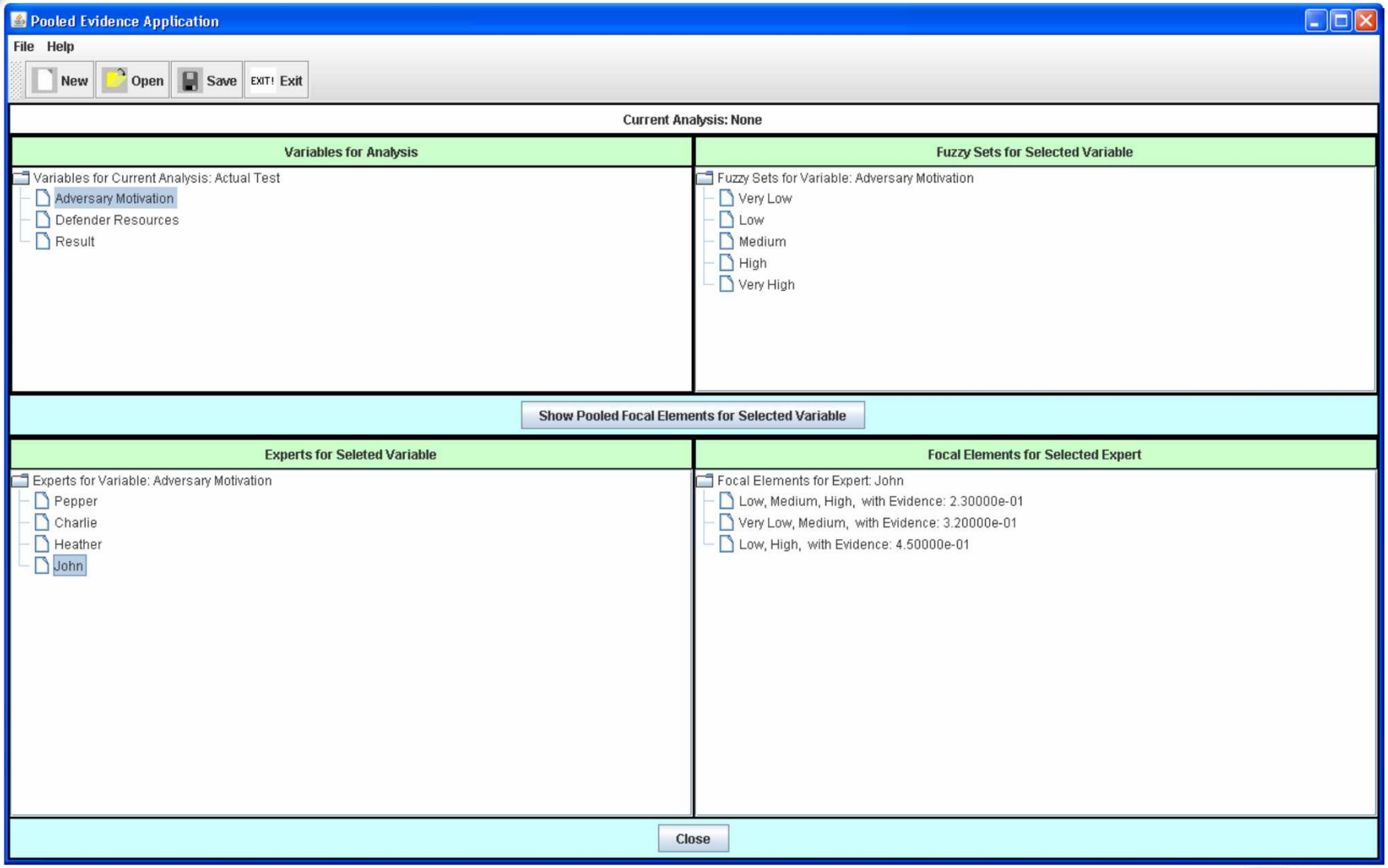
4 Experts Assign Evidence



jldarby@sandia.gov

Sandia National Laboratories

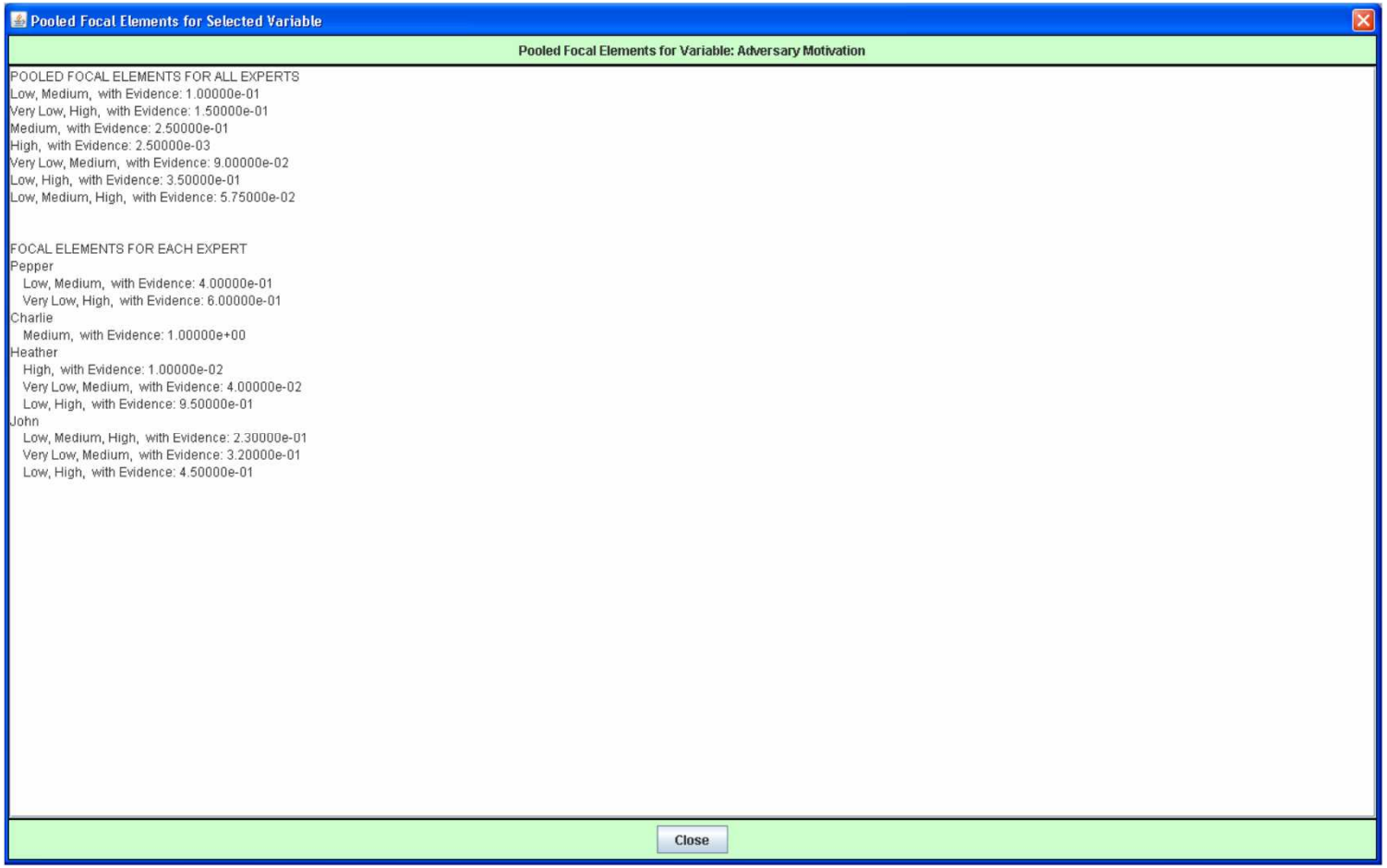
PoolEvidence: Example Application



jldarby@sandia.gov

Sandia National Laboratories

PoolEvidence



jldarby@sandia.gov

Sandia National Laboratories

Summary

- Uncertainty
  - Belief/Plausibility for Epistemic Uncertainty
- Qualitative Variables
  - Purely Linguistic Fuzzy Sets
- Combining Variables
  - Approximate Reasoning
  - Convolution with Belief/Plausibility
- Custom Software Tools
  - BeliefConvolution
  - LinguisticBelief
  - PoolEvidence

jldarby@sandia.gov

Sandia National Laboratories