Raytheon

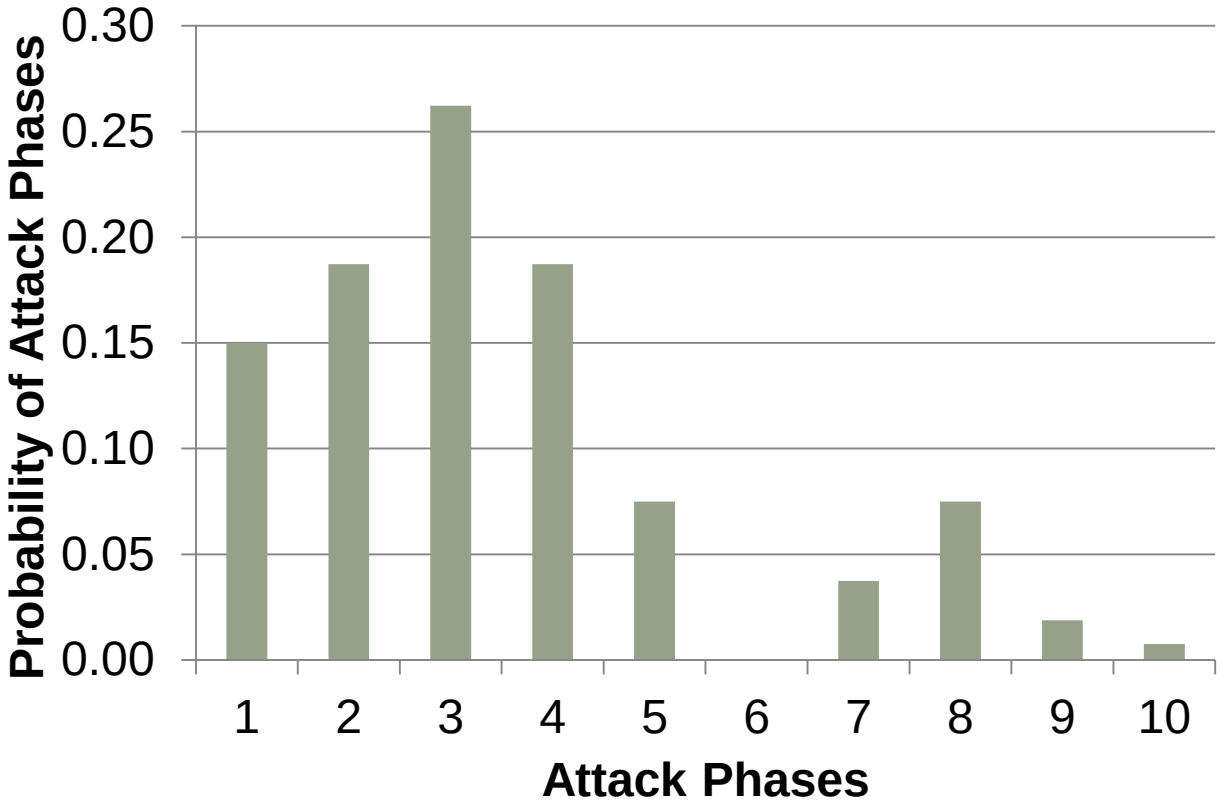
Customer Success Is Our Mission

Risk Metrics for Cyber Inference Assessment

Dr. Kenric P. Nelson
Raytheon Company
Sr. Principal Systems Engineer
November 12, 2014

What is the average uncertainty?

Given measurements regarding the phases of an attack, what is the average probability of the attack's progression?



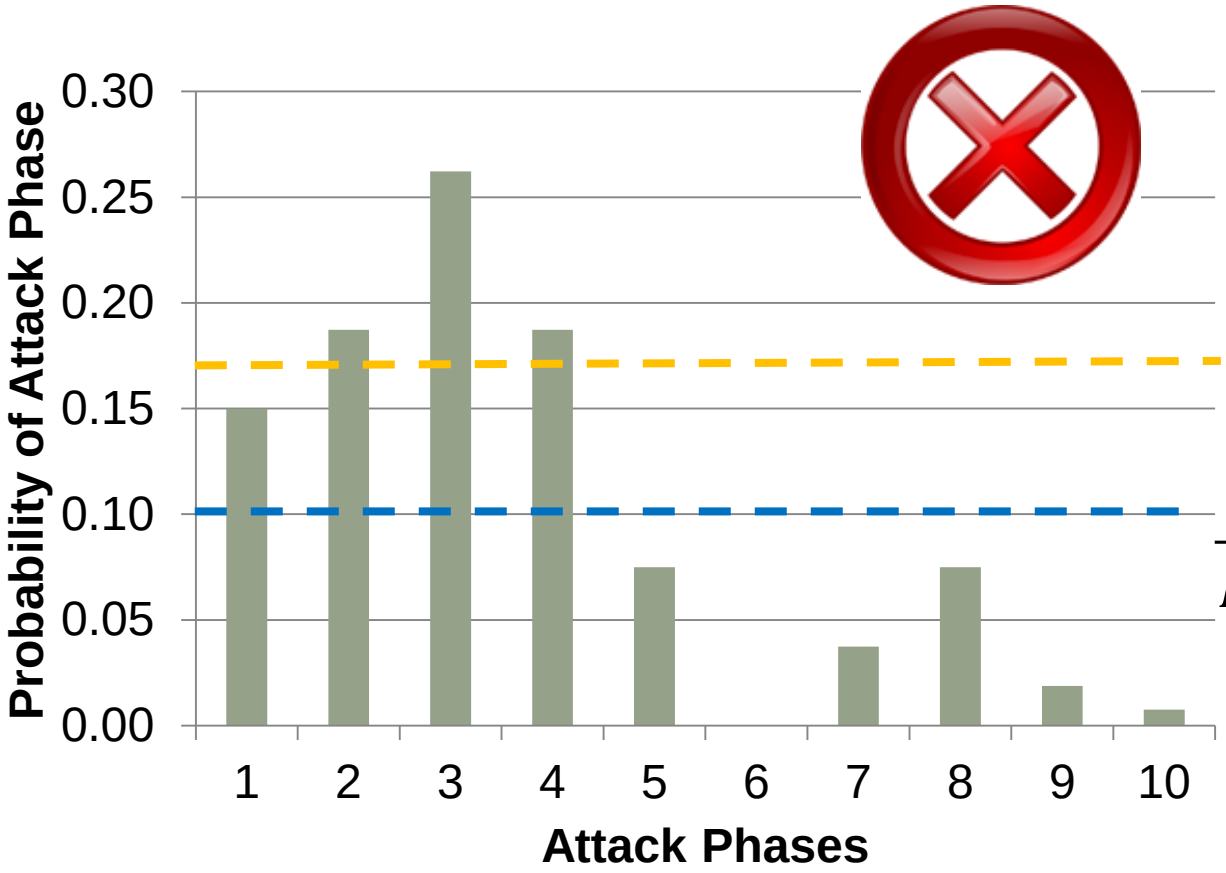
Attack phases might include scanning, enumeration, access, pilfering, etc.

Outline

- Average Uncertainty: Making info metrics intuitive
- Assessing threat models: Problems with Scoring Rules
  - Lack of clarity regarding which rules are appropriate
  - Information theoretic rule – logarithmic rule – is very sensitive
  - Results are unintuitive – what is entropy? How does it relate to uncertainty?
- The Risk Profile
  - Spectrum of algorithm performance relative to degree of risk tolerance
  - Originates from and encapsulates Tsallis entropy – information for nonlinear systems
  - Example analysis for classification systems
- Conclusion & Suggested Applications

What is the average uncertainty?
Arithmetic mean: seems intuitive but incorrect

Raytheon



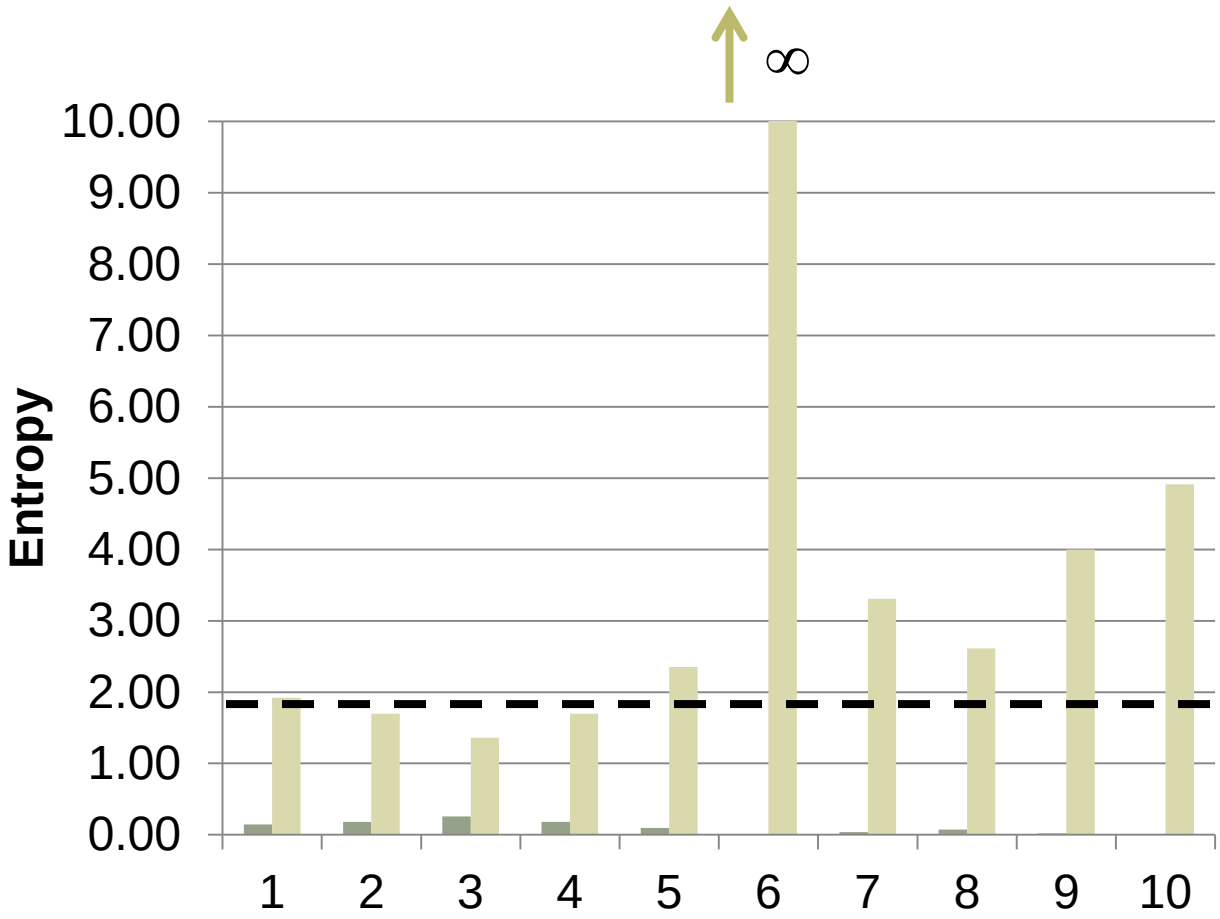
$$\sum_i p_i p_i = \sum_i p_i^2$$

$$\frac{1}{N}\sum_i p_i = \frac{1}{N}$$

Not the arithmetic mean;
Nor the weighted mean

What is the average uncertainty of threats? Information theory: accurate but unintuitive

Raytheon



$$-\ln(p_i)$$

$$-\sum_i p_i \ln p_i$$

Often interpreted as a length in natural bits (nats), but how does this relate to the original probabilities?

The average uncertainty: An intuitive approach to information theory

Raytheon

All information theoretic analysis can be translated from entropy to an average probability

Translation to probability scale is $e^{-\text{Entropy Function}}$

Info-Metric	Entropy Scale	Probability Scale
Entropy	$-\sum_i p_i \ln p_i$	$\prod_i (p_i)^{p_i}$
Divergence	$-\sum_i p_i \ln\left(q_i / p_i\right)$	$\prod_i \left(q_i / p_i\right)^{p_i}$
Cross-Entropy	$-\sum_i p_i \ln q_i$	$\prod_i (q_i)^{p_i}$

Information metrics as Probabilities

Info-Metric	Entropy Scale	Probability Scale
Entropy	$-\sum_i p_i \ln p_i$	$\prod_i (p_i)^{p_i}$
Divergence	$-\sum_i p_i \ln\left(q_i / p_i\right)$	$\prod_i \left(q_i / p_i\right)^{p_i}$
Cross-Entropy	$-\sum_i p_i \ln q_i$	$\prod_i (q_i)^{p_i}$

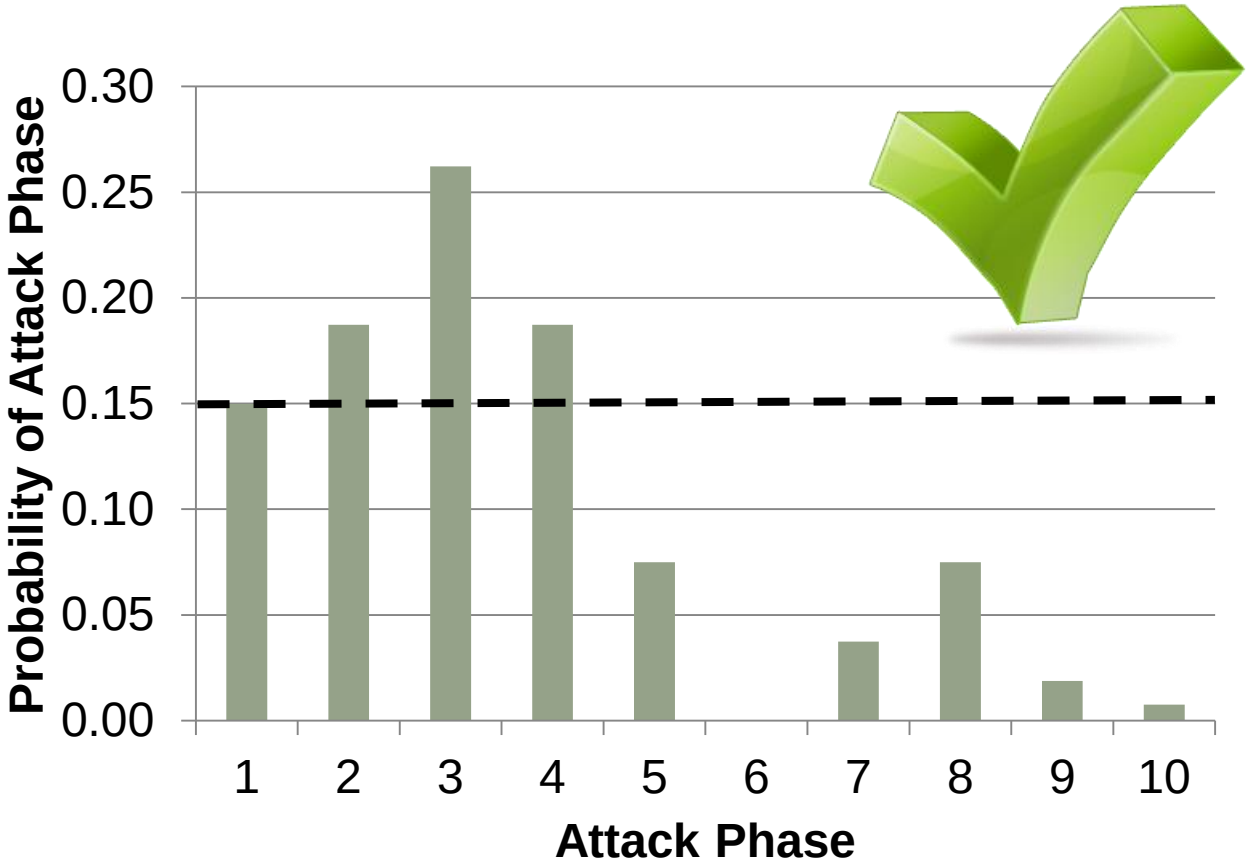
Information gain = reduction in Shannon entropy

Equivalently Shannon teaches the average probability

Information gain = increase in average probability

What is the average uncertainty of threats? The Weighted Geometric Mean !!

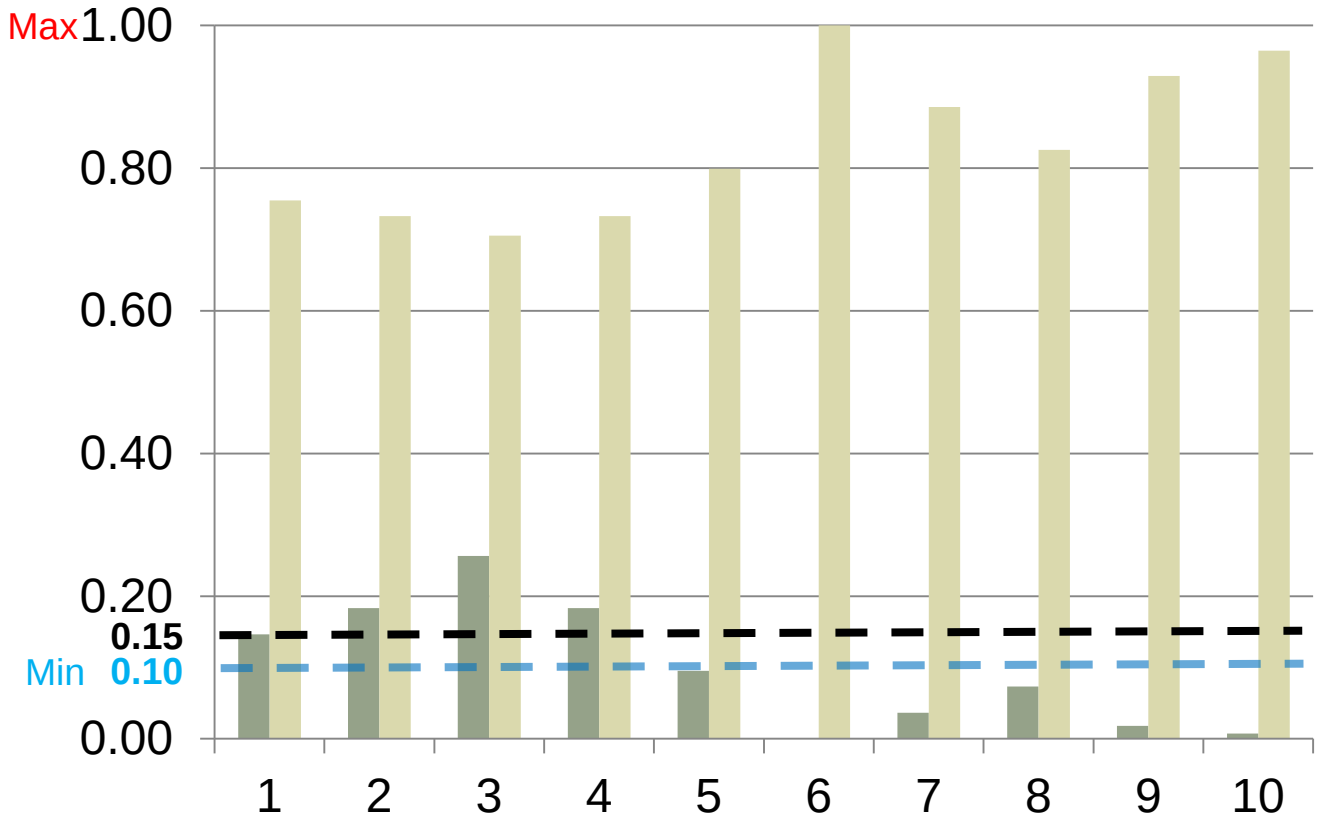
Raytheon



$$\prod_i p_i^{p_i}$$

Power and accuracy of information theory
Simplicity & intuition of average probability

Interpreting $p_i^{p_i}$



$$p_i^{p_i}$$

$$\prod_i p_i^{p_i}$$

Represents probability of each event p_i occurring p_i times

Product is all events occurring a total of once; i.e. average

Accuracy of threat Assessment?

- Purpose is to assess the accuracy of probabilistic forecasts
- Comparison between two distributions:
  - Distribution of forecasts produced by algorithms, models, & analysts
  - Distribution of test data used to evaluate the performance of analysts
- Well established performance metrics based on decision boundaries
  - Confusion Matrix – percent correct classification & percent of decision errors
  - Receiver Operator Curve – how does decision boundary affect confusion matrix
- Accuracy of probabilistic forecasts much harder to assess
  - Again, arithmetic mean of true event probabilities is not correct
  - Instead a scoring rule needed which weights the value of a probability; this value can be averaged
  - Information theory: value of probability is negative logarithm, but oversensitive
  - Most popular alternative: Mean-square average of the reported probabilities
  - Countless alternatives: starting with any concave utility function, can derive a "Proper Scoring Rule" which encourage honesty in the mean, but modifies the risk associated with variation in the forecast
- Demonstrate approach which uses a risk-biased info metric

Coupled surprisal modifies info metric

Nonlinear metric:

$$\ln_{-\kappa} 1/p = -\ln_{\kappa} p \equiv -\frac{p^{\kappa_{mult}} - 1}{\kappa_{add}}$$

Graph shows $\kappa_{add} = \kappa_{mult}$

$$-\frac{d}{dp}\ln_{\kappa} p = 1$$

$$\text{If } \kappa_{add} = \frac{\kappa_{mult}}{1+\kappa_{mult}}$$

$$\text{Then } -\int_0^1 \ln_{\kappa} p = 1$$

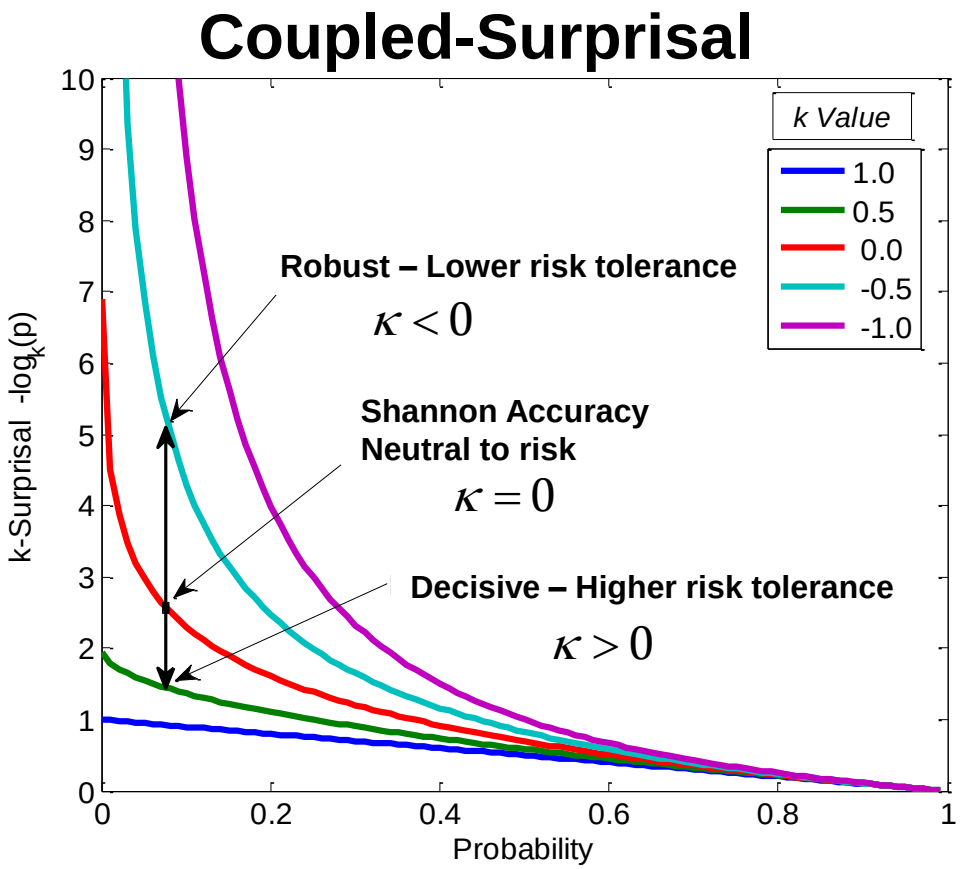
Coupled Entropy:

$$-\sum_i p_i \ln_{\kappa} p_i$$

This is the dual $q^* = 2 - q$

Tsallis entropy $\kappa^* = -\kappa$

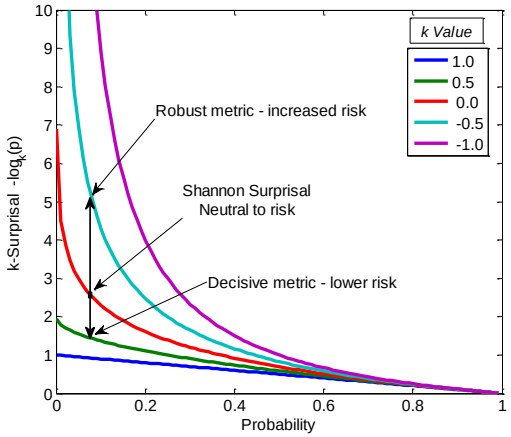
- Properties of coupled surprisal
  - Defined by deformation from additive metric
  - Related to the degree of risk tolerance



Coupled-Surprisal ⇒ Gen. Mean

- Utilize just the coupled-surprisal to form Risk Profile
- Average coupled-surprisal is biased, local score

Coupled-Suprisal $\xrightarrow[Arithmetic\ Average]{\Rightarrow}$ Coupled Cross-Entropy $\xrightarrow[Coupled-Exp]{\Rightarrow}$ Generalized Mean



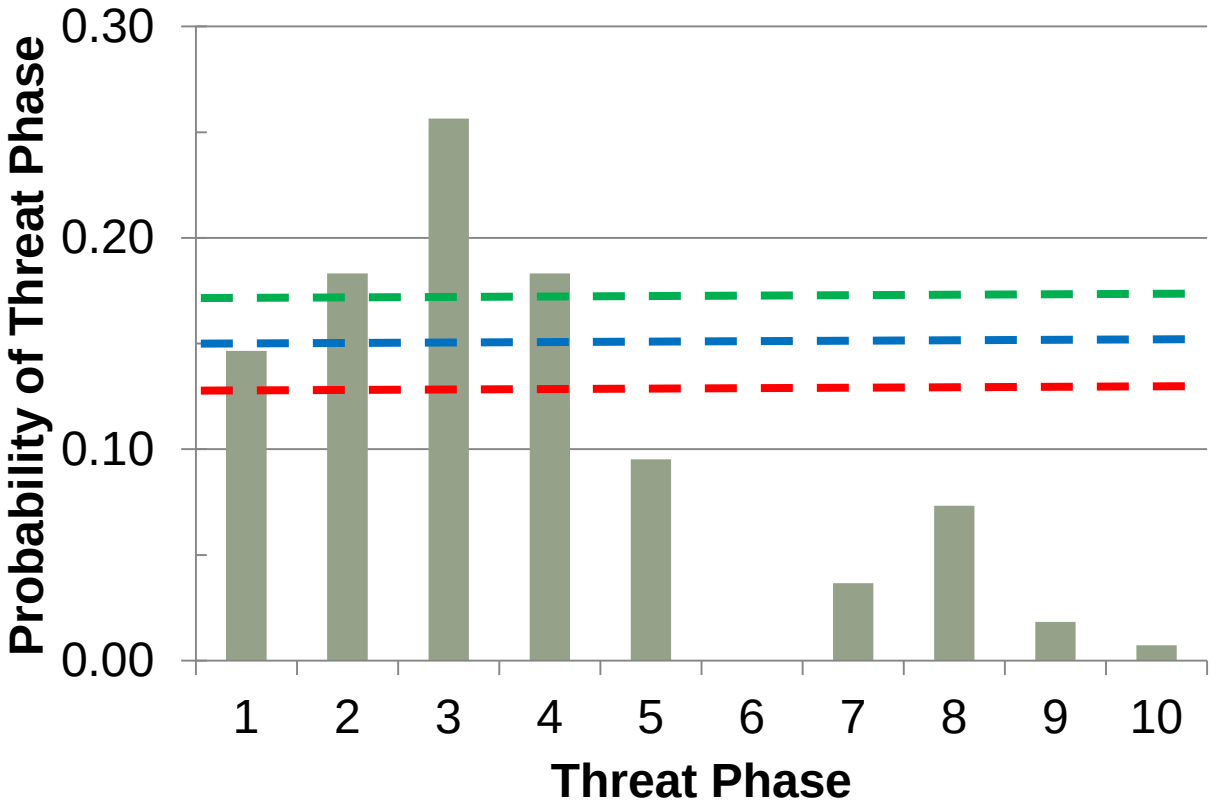
$$P_{avg}^{(\kappa)}(\mathbf{p},\mathbf{q}\,|\,\mathbf{x}_{truth}) = \left(\frac{1}{N}\sum_{i=1}^{N} q_{i,truth}^{\kappa} \right)^{\frac{1}{\kappa}}$$

- (κ >0) Decisive – finite cost
- (κ <0) Robust – infinite cost

- Shannon Entropy ($\kappa = 0$): Log average → Geometric Mean
- Generalized mean can also be derived from Renyi Entropy

Illustration of bounds using generalized mean

Raytheon



$$\left(\sum_{i=1}^{10} p_i^{1+\kappa}\right)^{1/\kappa}$$

$\kappa = 1,\ p = 0.17$

$\kappa = 0,\ p = 0.15$

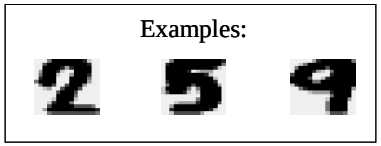
$\kappa = {}^{-2}/_{3},\ p = 0.13$

$$\kappa_{robust} = \frac{-2\kappa_{decisive}}{2+\kappa_{decisive}}$$

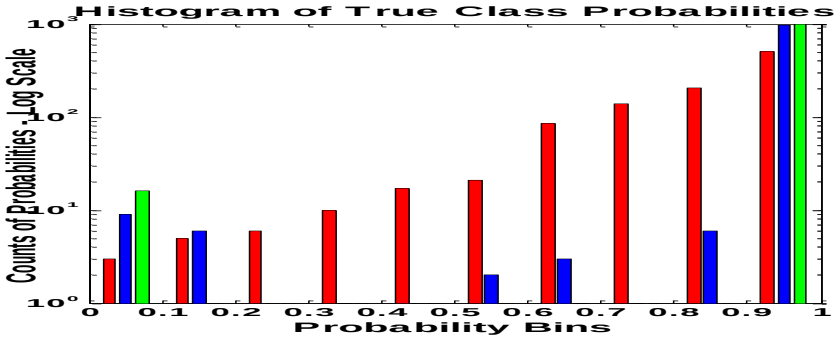
The Risk Profile: A scoring rule based on the degree of risk tolerance

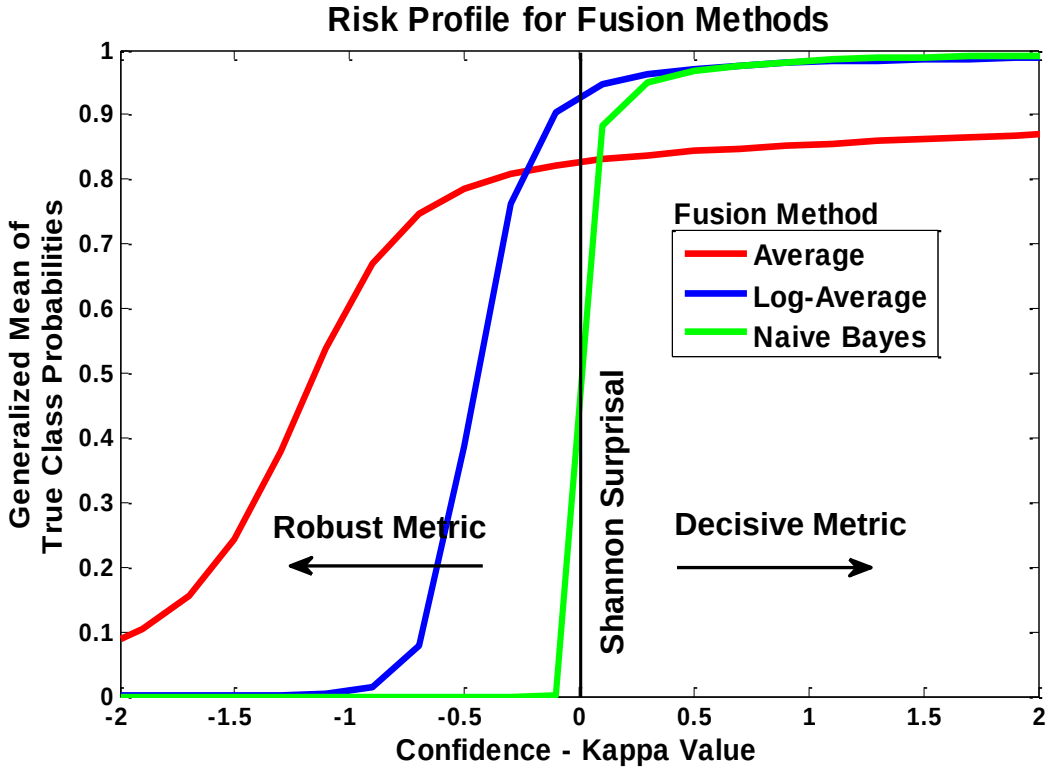
Raytheon

- Input is the histogram of true state probabilities
- Output is spectrum of performance versus risk tolerance
- Provides insight into forecasts:
  - **Decisiveness**
  - **Accuracy**
  - **Robustness**
- Example



Fusion of Image Features



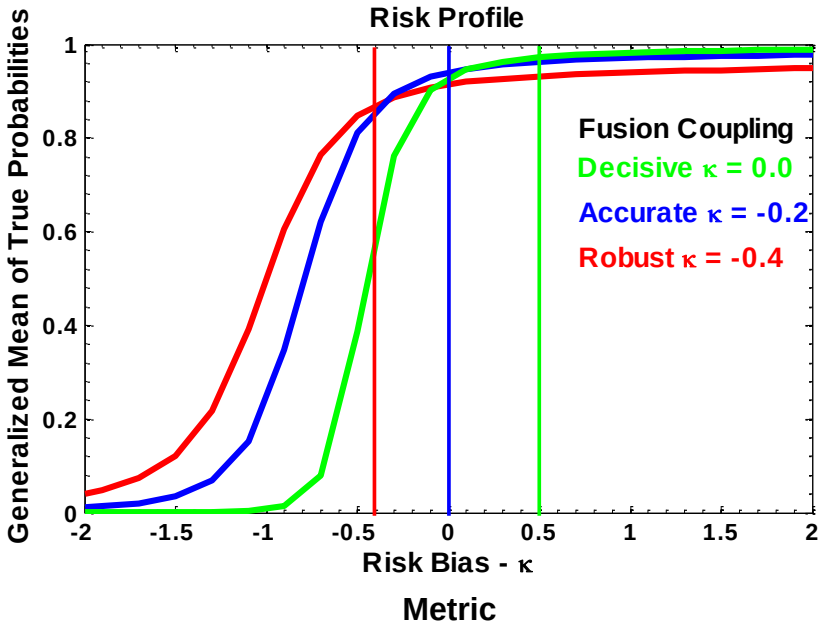


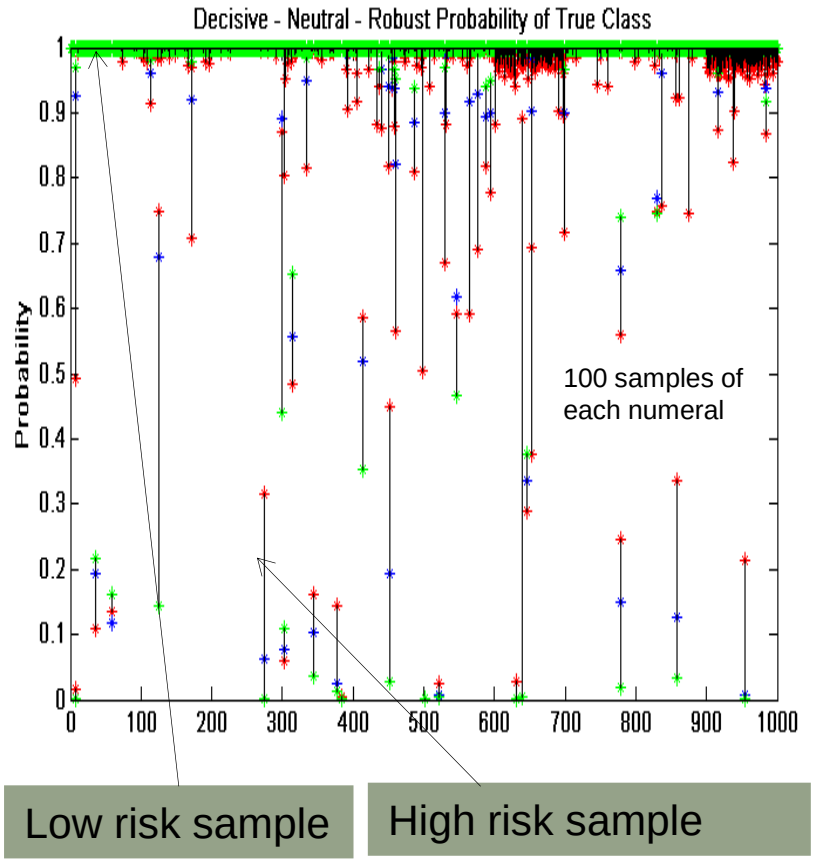
Fusion & Info-Metric use Generalized Mean

Raytheon

Examples:

Fusion with Generalized Mean of 6 image features
Correct Classification 98%
Distribution of Probabilities Modified by Risk Bias





Using risk bias for bounds rather than variance

US Patent # US8595177 B1

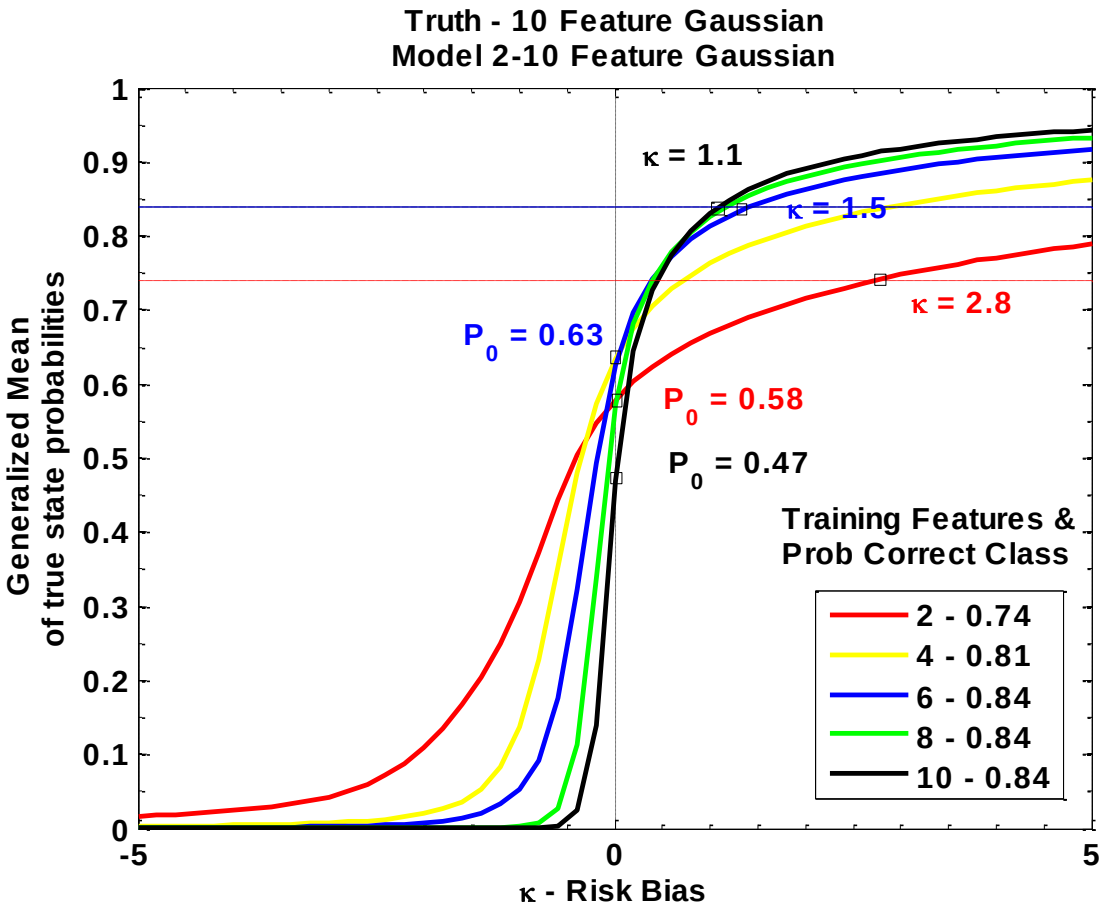
Overfitting high-dimensional models

Truth & Model

- Data generated from 10-D Independent Gaussian
- Training data estimates μ & σ
- Model is Gaussian
- Model has 2-10 Dim.

Results

- Decision Accuracy plateaus at 6 features
- Probability Accuracy degrades from
  - 0.63 with 6 features
  - to 0.47 with 10 features



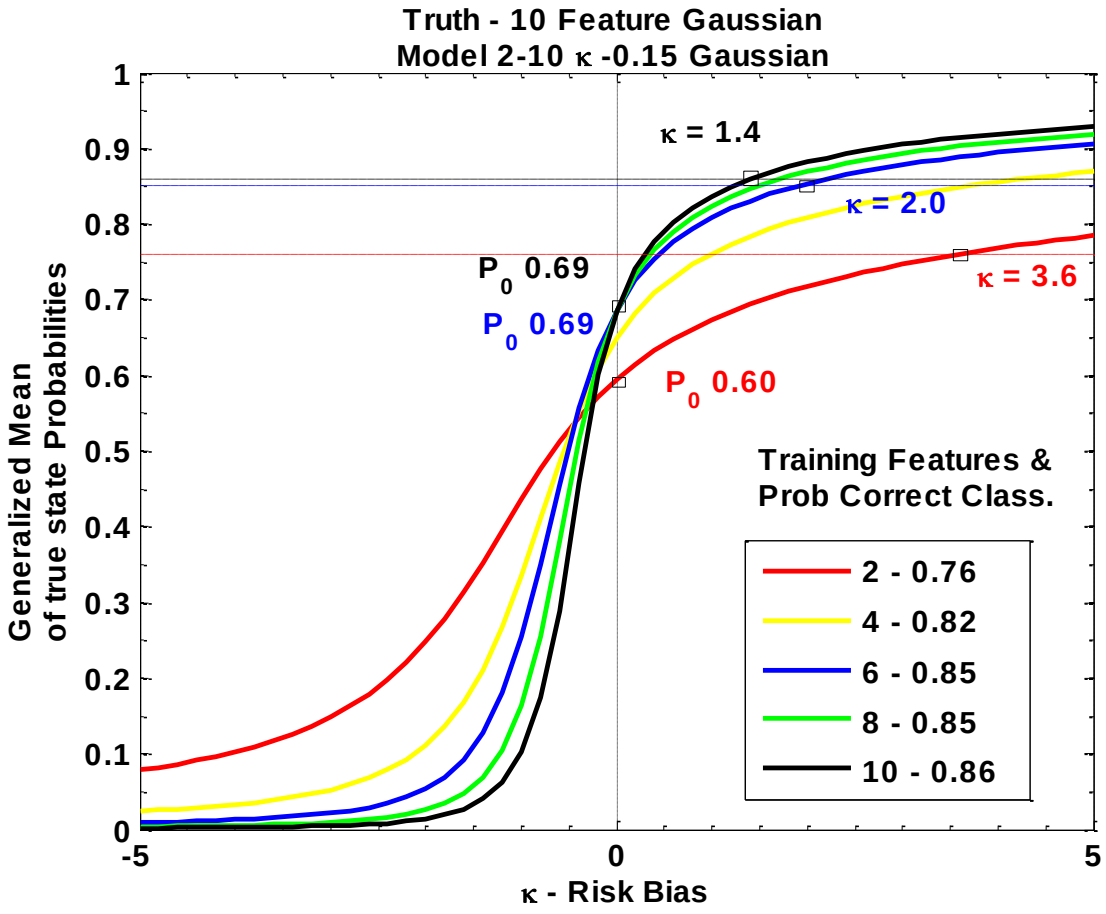
Robust Heavy-tail Model

Truth & Model

- Data generated from 10-D Independent Gaussian
- Training data estimates μ & σ
- Model is Heavy-Tail – robust against outliers
- Model has 2-10 Dim.

Results

- Decision Accuracy improves to 0.86 at Dim = 10
- Probability Accuracy
  - stable at 0.86 for dim > 6



Conclusion

- Average uncertainty is the **Geometric Mean** of probabilities
- Risk assessment of forecasting algorithms requires …
  - Decisiveness: is there enough certainty to make good decisions?
  - Accuracy: are the probabilistic forecasts honest about the uncertainty?
  - Robustness: how sensitive is the algorithm to the testing data?
- Average risk-biased uncertainty is the **Generalized Mean**
- Resulting analytical tool is the **Risk Profile**
  - Information theoretic measure of algorithm performance versus risk
  - Uses the familiar probability scale so results are intuitive
  - Spectrum of performance provides rich insight into characteristics of algorithms
- Application to Cyber Metrics
  - Evaluation of tools used to forecast threats
  - Provides insight about how well an algorithm is balancing forecasting risks