U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT COMMAND ARMY RESEARCH LABORATORY

Dealing with Epistemic Uncertainty in Information Fusion Systems

Lance M Kaplan (DEVCOM ARL)

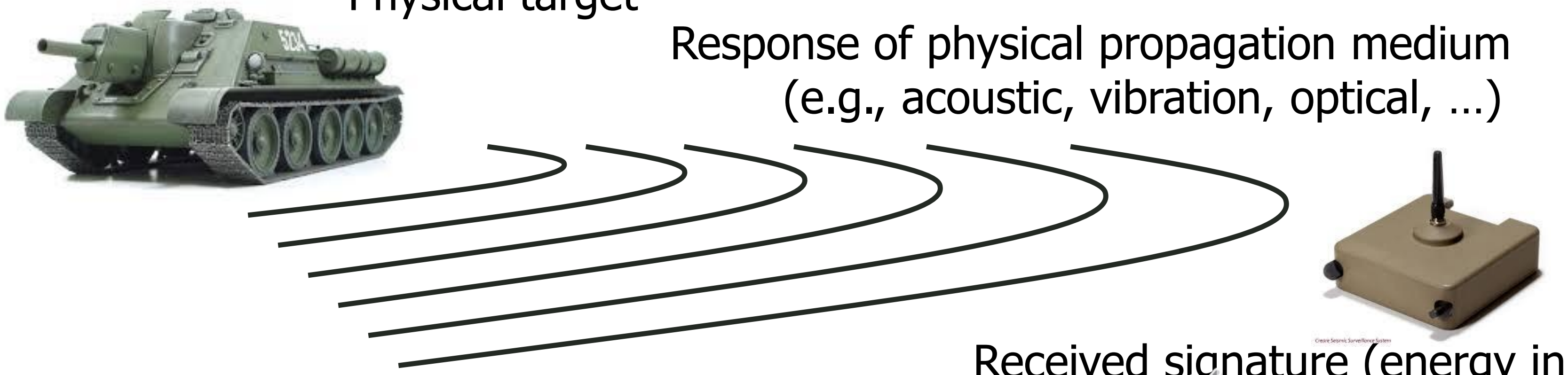
Collaborators: Magdalena Ivanovska (UiO), Federico Cerutti (Univ. Brescia), Murat Sensoy (Amazon), Kumar Vijay Mishra (ARL), Conrad Hougen (U Mich), James Z. Hare (ARL), Cesar Uribe (Rice), Ali Jadbabaie (MIT), Venugopal V. Veeravalli (UIUC)

13 SEP 2022

ALEATORIC UNCERTAINTY

Physical target

Response of physical propagation medium
(e.g., acoustic, vibration, optical, …)

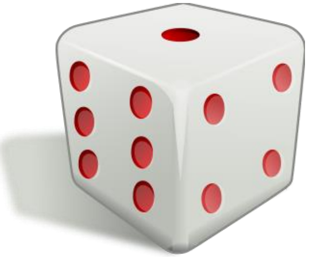


Received signature (energy in
multiple signal frequency bands)

Physical event

Random component of sensor measurements

Likelihood	$f(x\|\theta)$	Parameters θ	Entropy
Gaussian	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	Mean μ Precision $\tau = \frac{1}{\sigma^2}$	$\frac{1}{2} + \frac{1}{2}\log(2\pi\sigma^2)$
Multinomial	$\prod_i (p_i)^{\delta_{x,i}}$	Categorical probabilities p_i	$-\sum_i p_i \log p_i$



Random component of event prediction

EPISTEMIC UNCERTAINTY

Given observations $X_{1:N} = \{x_1, \dots, x_N\}$,
what do we know about the parameters θ?

Full information depicts parameter uncertainty

Likelihood	Conjugate Prior	$f(\theta \mid X_{1:N})$
Gaussian	Gaussian Inverse Gamma	$f(\mu,\tau) = \sqrt{\frac{\lambda}{2\pi}} \left(\frac{S}{2}\right)^{\alpha} \frac{\tau^{\alpha-\frac{1}{2}} e^{-\frac{\tau\lambda}{2}(\mu-m)^2} e^{-\frac{S\tau}{2}}}{\Gamma(\alpha)}$
Multinomial	Dirichlet	$f(\boldsymbol{p}) = \frac{1}{\mathrm{B}(\boldsymbol{\alpha})} \prod_i (p_i)^{\alpha_i - 1}$

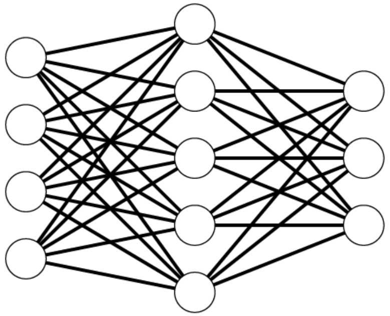
What are the parameters of the aleatoric model?

No expression of uncertainty

MAP estimate: $\hat{\theta} = \mathrm{argmax}_{\theta}\, f(\theta \mid X_{1:N})$

UNCERTAINTY FOR AI & ML REASONING

Perception
of simple events



Detection of complex
events with fixed logic

$$\frac{\Delta \to \Lambda, C \qquad C, \Gamma \to \Theta}{\Delta, \Gamma \to \Lambda, \Theta}$$

$$\frac{\Gamma \to \Theta}{\Gamma \to \Theta, C}$$

Predicted Complex Event

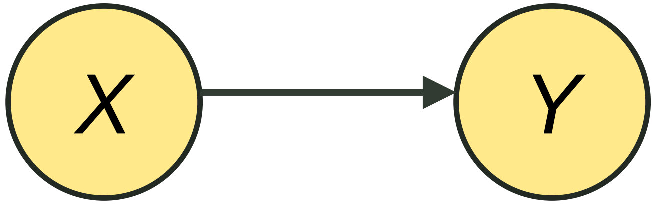
Neuro-Symbolic architecture facilitates:

- **Learning *complex events* from sparse data**
- **Tellibility – enables a domain expert to inject scientifically grounded reasoning rules**
- **Explainability – provides a train of reasoning step for a human decision maker**
- **Adaptability – to adjust to new environments**

How does **uncertainty** percolate through the neuro- and symbolic layers?

UNCERTAINTY – SYMBOLIC REASONING – SIMPLE EXAMPLE

$$p(X,Y) = p(Y|X)p(Y)$$



Parameters

$$\theta_x \rightarrow p(X=1)$$
$$\theta_{y|x} \rightarrow p(y=1|x=1)$$
$$\theta_{y|\bar{x}} \rightarrow p(y=1|x=0)$$

Ins	*X*	*Y*	Likelihood
1	0	1	$\theta_{y\|\bar{x}}(1-\theta_x)$
2	0	0	$(1-\theta_{y\|\bar{x}})(1-\theta_x)$
3	1	1	$\theta_{y\|x}\theta_x$
4	0	0	$(1-\theta_{y\|\bar{x}})(1-\theta_x)$
5	1	1	$\theta_{y\|x}\theta_x$
6	1	0	$(1-\theta_{y\|x})\theta_x$
7	0	1	$\theta_{y\|\bar{x}}(1-\theta_x)$
8	1	0	$(1-\theta_{y\|x})\theta_x$
9	0	0	$(1-\theta_{y\|\bar{x}})(1-\theta_x)$
⋮	⋮	⋮	⋮

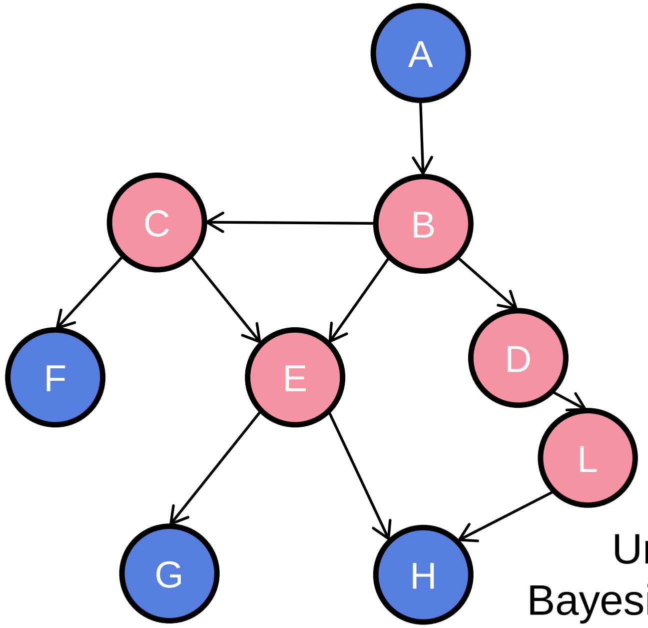
$$f(\theta_x, \theta_{y|x}, \theta_{y|\bar{x}}) \propto \theta_x^{n_{yx}+n_{\bar{y}x}}(1-\theta_x)^{n_{y\bar{x}}+n_{\bar{y}\bar{x}}}\theta_{y|x}^{n_{yx}}(1-\theta_{y|x})^{n_{\bar{y}x}}\theta_{y|\bar{x}}^{n_{y\bar{x}}}(1-\theta_{y|\bar{x}})^{n_{\bar{y}\bar{x}}}$$

Posterior

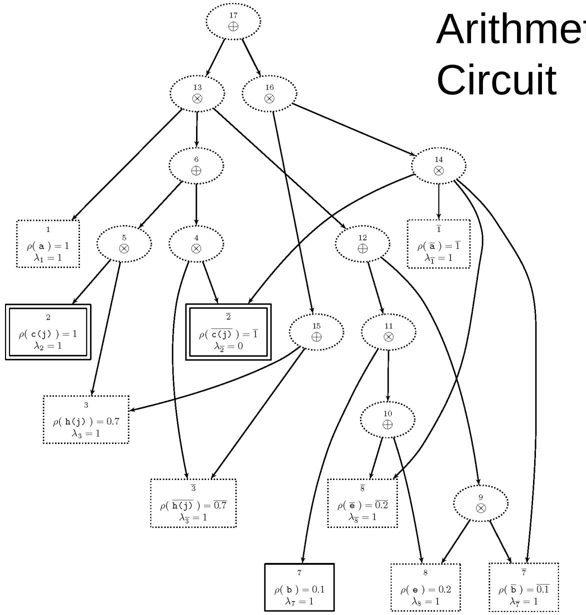
UNCERTAINTY – SYMBOLIC REASONING

```
ω2::burglary.
ω3::earthquake.
ω4::hears_alarm(john).
alarm :- burglary.
alarm :- earthquake.
calls(john) :- alarm, hears_alarm(john).
evidence(calls(john)).
query(burglary).
```

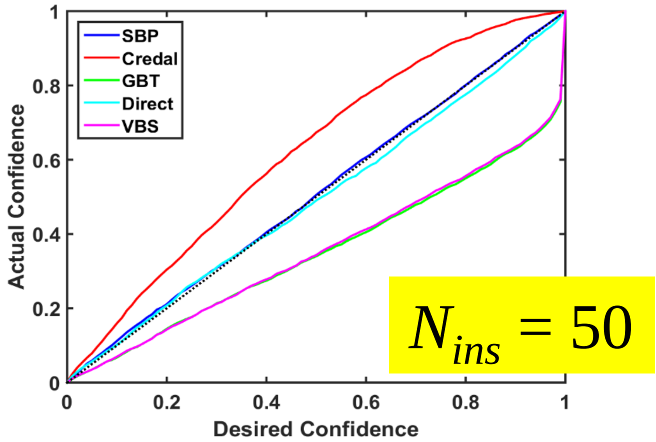
Beta ProbLog



Uncertain Bayesian Networks



Arithmetic Circuit



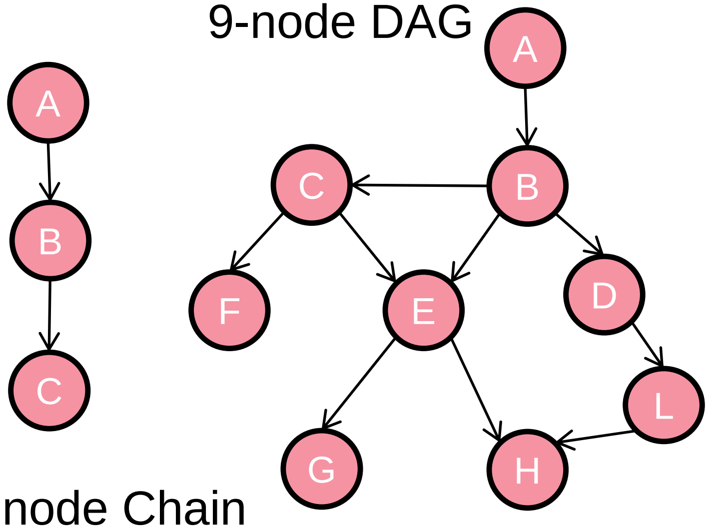
Desired Confidence Bound Divergence

Second-Order Learning and Inferencing via the δ-method

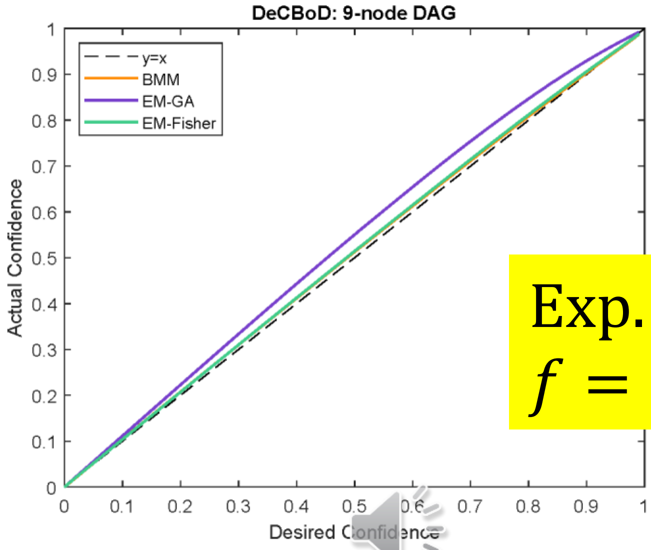
https://github.com/federicocerutti/CPB

LEARNING WITH INCOMPLETE TRAINING DATA

9-node DAG



3-node Chain



Exp. A
$f = 0.5$

Generated 1000 Certain Bayesian Networks per structure

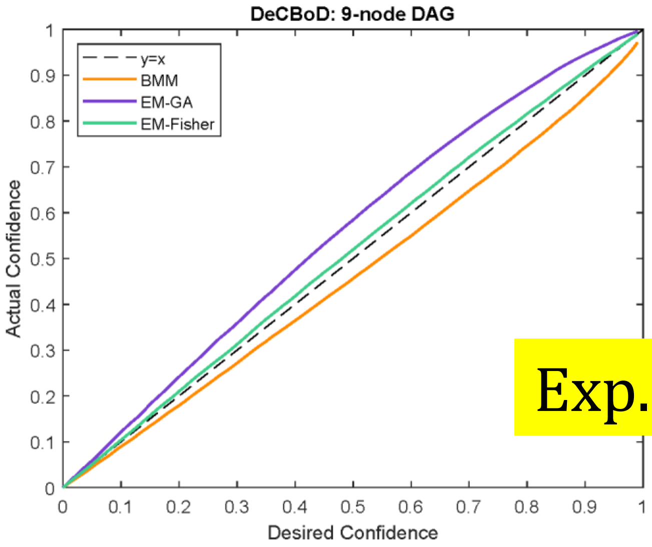
Trained one uncertain Bayesian networks with $N_{ins} = 120$ for each certain network

During inference, the assignment of observed variables and their values is random.

$|\mathbb{X}| = 3$

Experiment A: Each variable value is observed during training with probability of f.

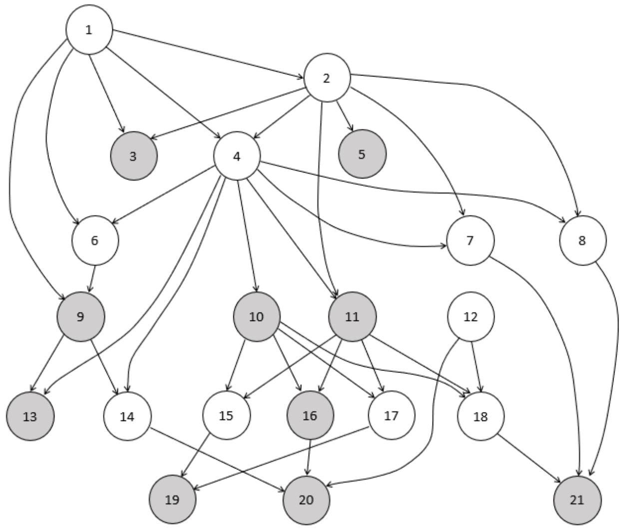
Three Node Chain - Mean Absolute DeCBoD									
f	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
BMM	0.0498	0.0270	0.0194	0.0083	0.0025	0.0045	0.0025	0.0030	0.0016
EM-GA	0.0406	0.0220	0.0242	0.0197	0.0119	0.0117	0.0138	0.0119	0.0132
EM-Fisher	0.0386	0.0182	0.0195	0.0136	0.0048	0.0046	0.0060	0.0038	0.0046
Nine Node DAG - Mean Absolute DeCBoD									
BMM	0.0635	0.0436	0.0252	0.0133	0.0083	0.0025	0.0017	0.0019	0.0017
EM-GA	0.0538	0.0396	0.0346	0.0331	0.0349	0.0356	0.0395	0.0437	0.0476
EM-Fisher	0.0487	0.0302	0.0202	0.0131	0.0096	0.0051	0.0041	0.0037	0.0036

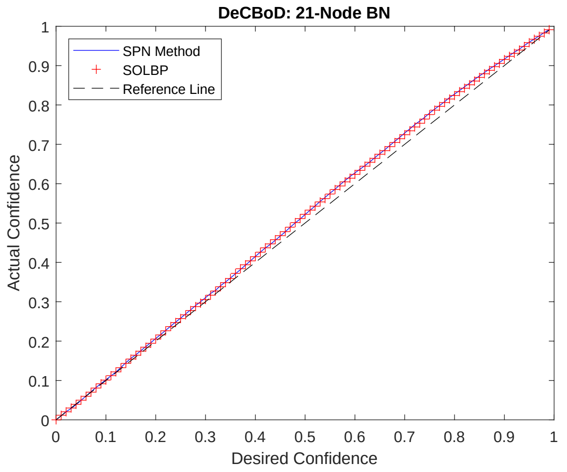


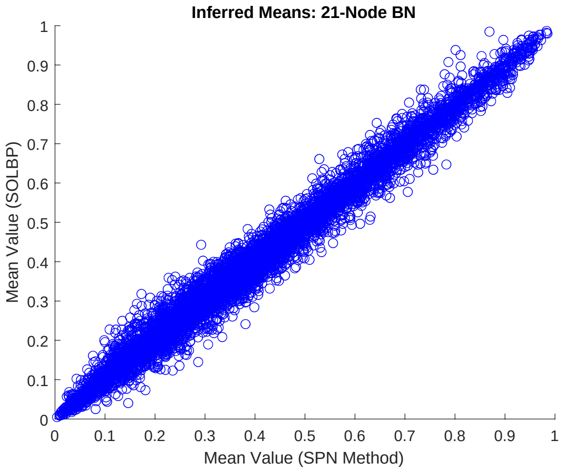
Exp. B

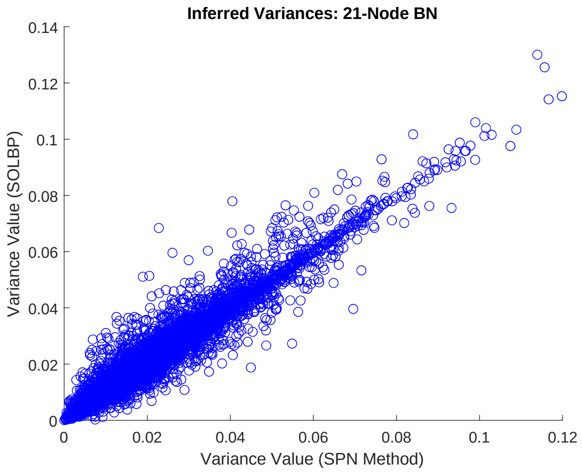
Experiment B: First 20 instantiations are complete and only the leaf variables are observed for the final 100 instantiations.

SECOND-ORDER LOOPY BELIEF PROPAGATION

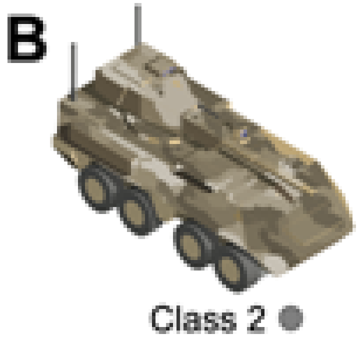




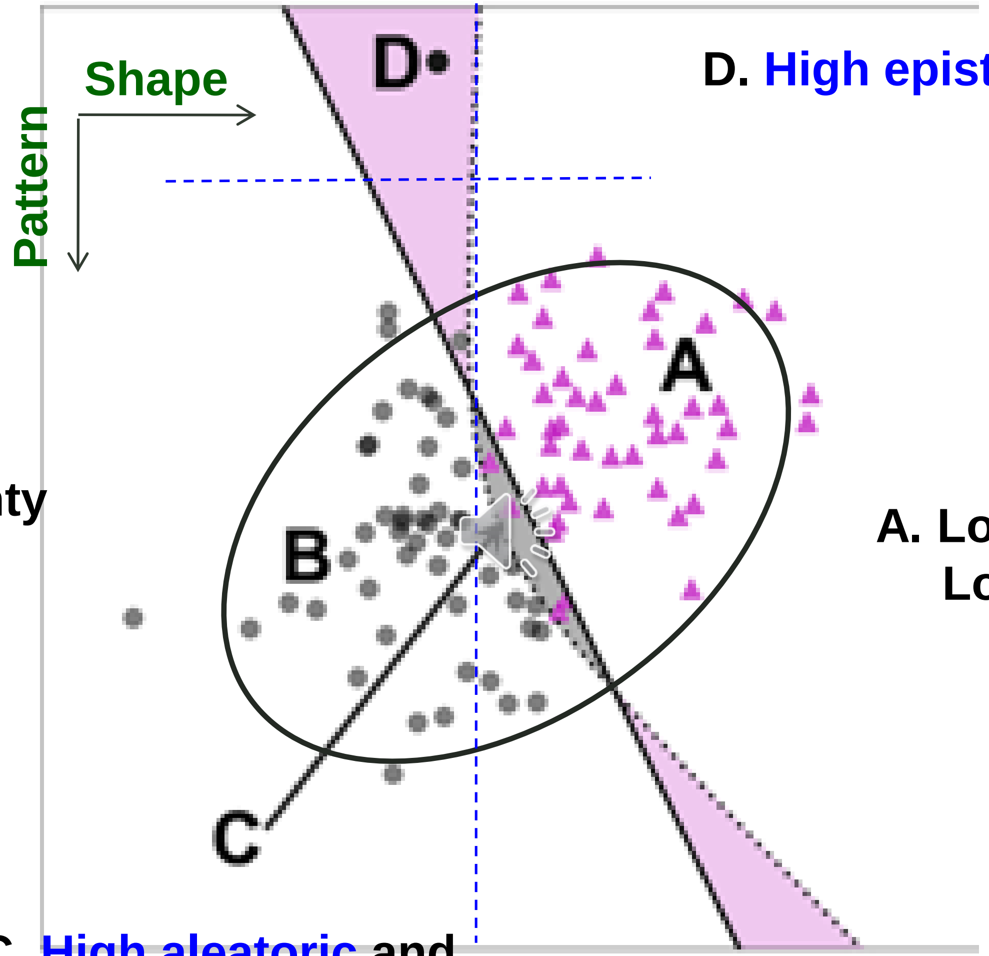




UNCERTAINTY IN MACHINE LEARNING – AN EVIDENTIAL VIEW

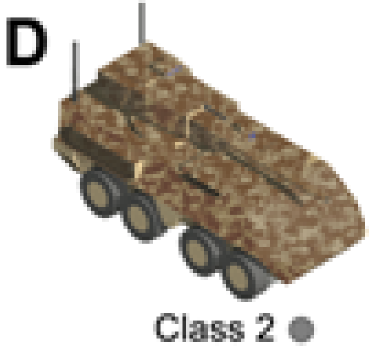


B. Low aleatoric and
Low epistemic uncertainty

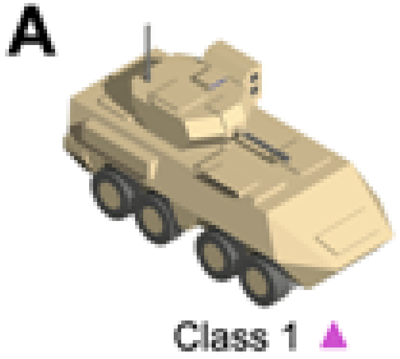


C. High aleatoric and
Low epistemic uncertainty

D. High epistemic uncertainty



A. Low aleatoric and
Low epistemic uncertainty

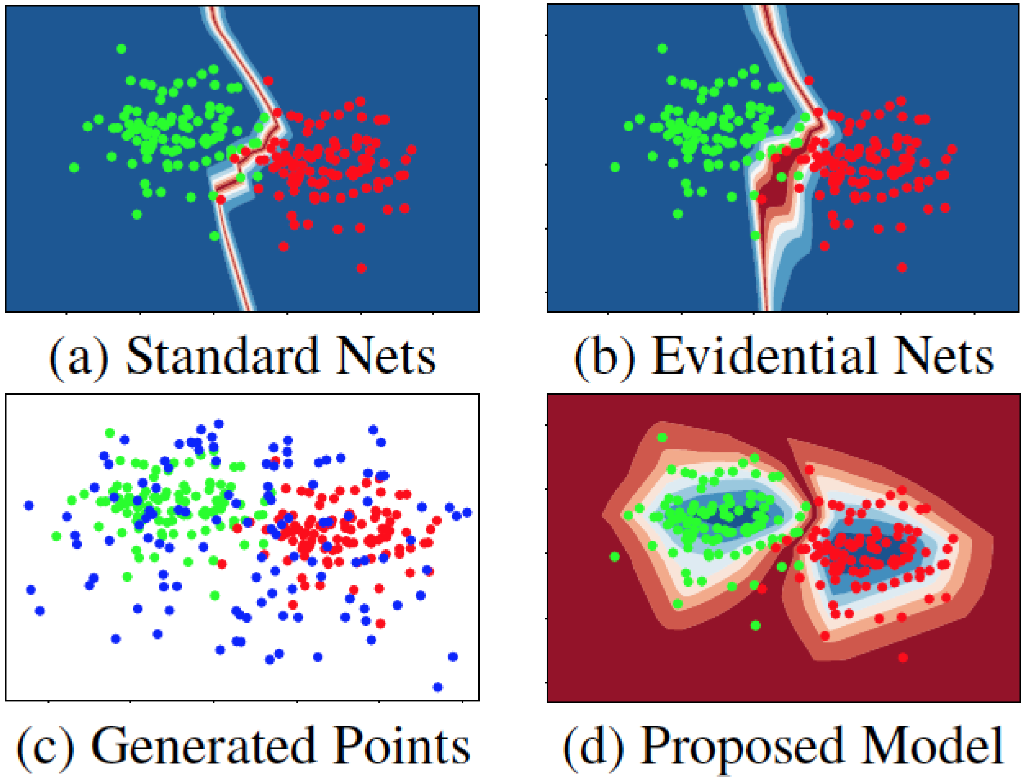


Epistemic uncertainty =systematic uncertainty
Aleatoric uncertainty = statistical uncertainty

$f(x;\theta)$ outputs a Dirichlet distribution representative of relevant evidence in the training data

GENERATIVE EVIDENTIAL DEEP LEARNING

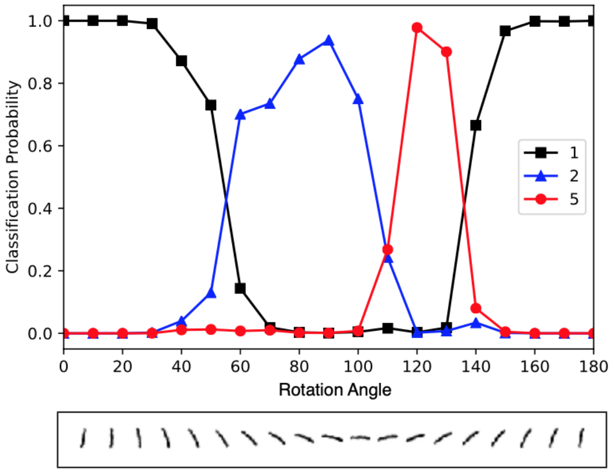
- Final layer of Neural Network represents evidence
- Noise-Contrastive Estimation (NCE) to learn the evidence
- Deep generative models combined with variational autoencoders to learn the noise distribution
- Leads to efficient quantification of epistemic and aleatoric uncertainty for deep classifiers.



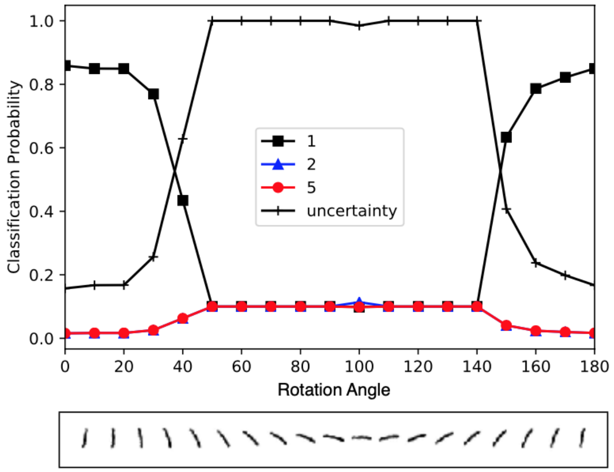
$$\mathcal{L}_1(\theta) = -\sum_{k=1}^{K} \Big[\mathop{\mathbb{E}}_{P_k(\boldsymbol{x})} [log(\sigma(f_k(\boldsymbol{x}|\theta)))] + \mathop{\mathbb{E}}_{P_{out}(\boldsymbol{x})} [log(1 - \sigma(f_k(\boldsymbol{x}|\theta)))] \Big]$$

$$\mathcal{L}_2(\theta|\boldsymbol{x}) = \beta \mathbb{KL}[D(\boldsymbol{p}_{-k}|\boldsymbol{\alpha}_{-k}) \;||\; D(\boldsymbol{p}_{-k}|\boldsymbol{1})]$$

Cross-entropy loss



EDL loss

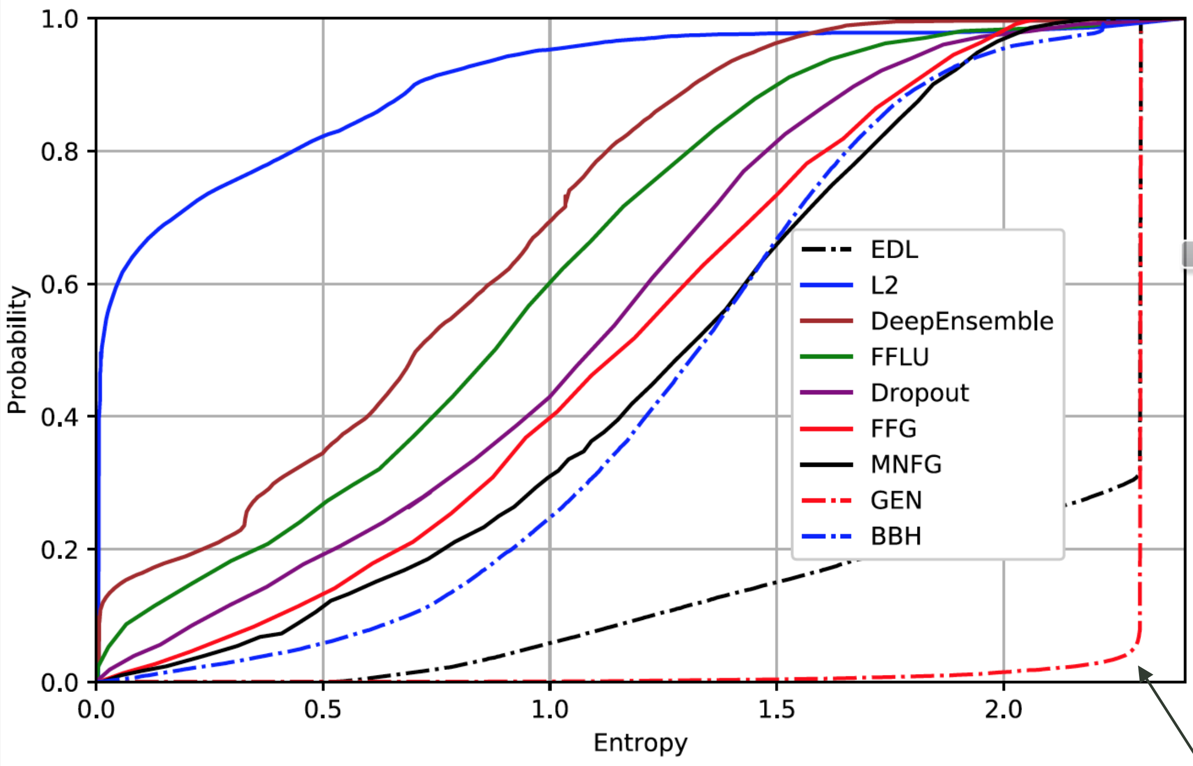


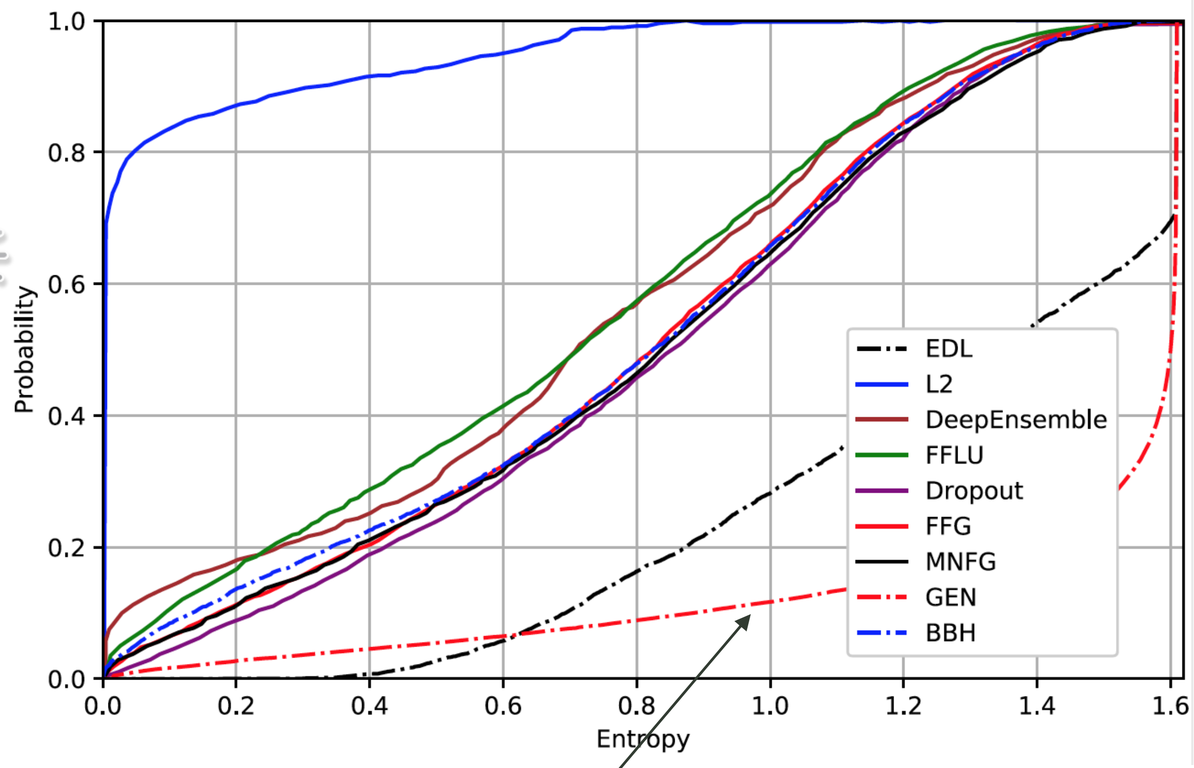
https://github.com/muratsensoy/muratsensoy.github.io

OUT OF DISTRIBUTION SAMPLES

notMNIST dataset

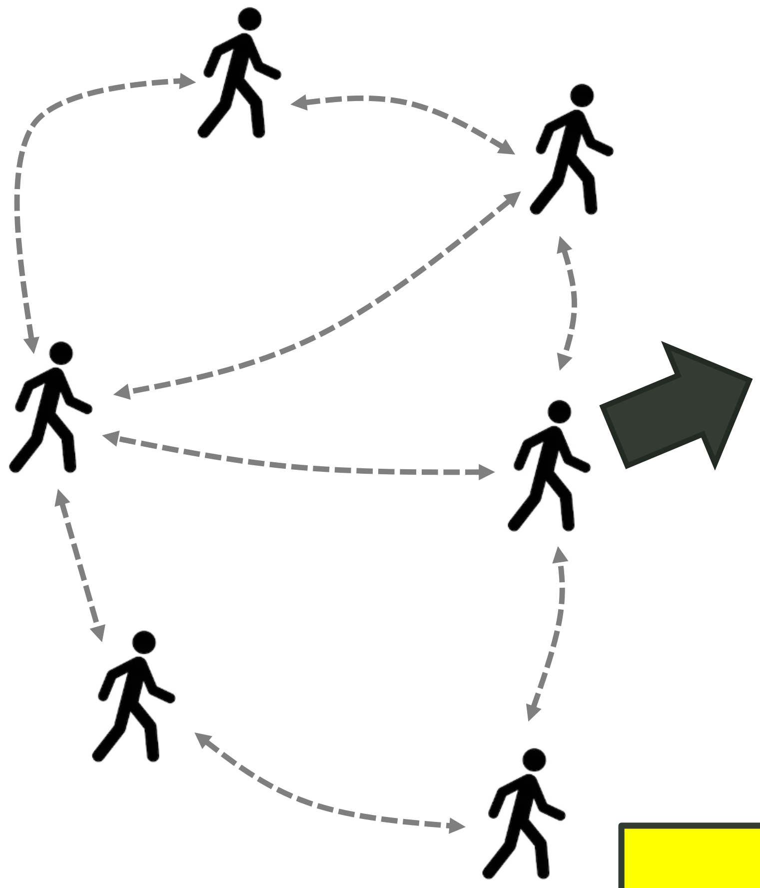
Last 5 categories from CIFAR10

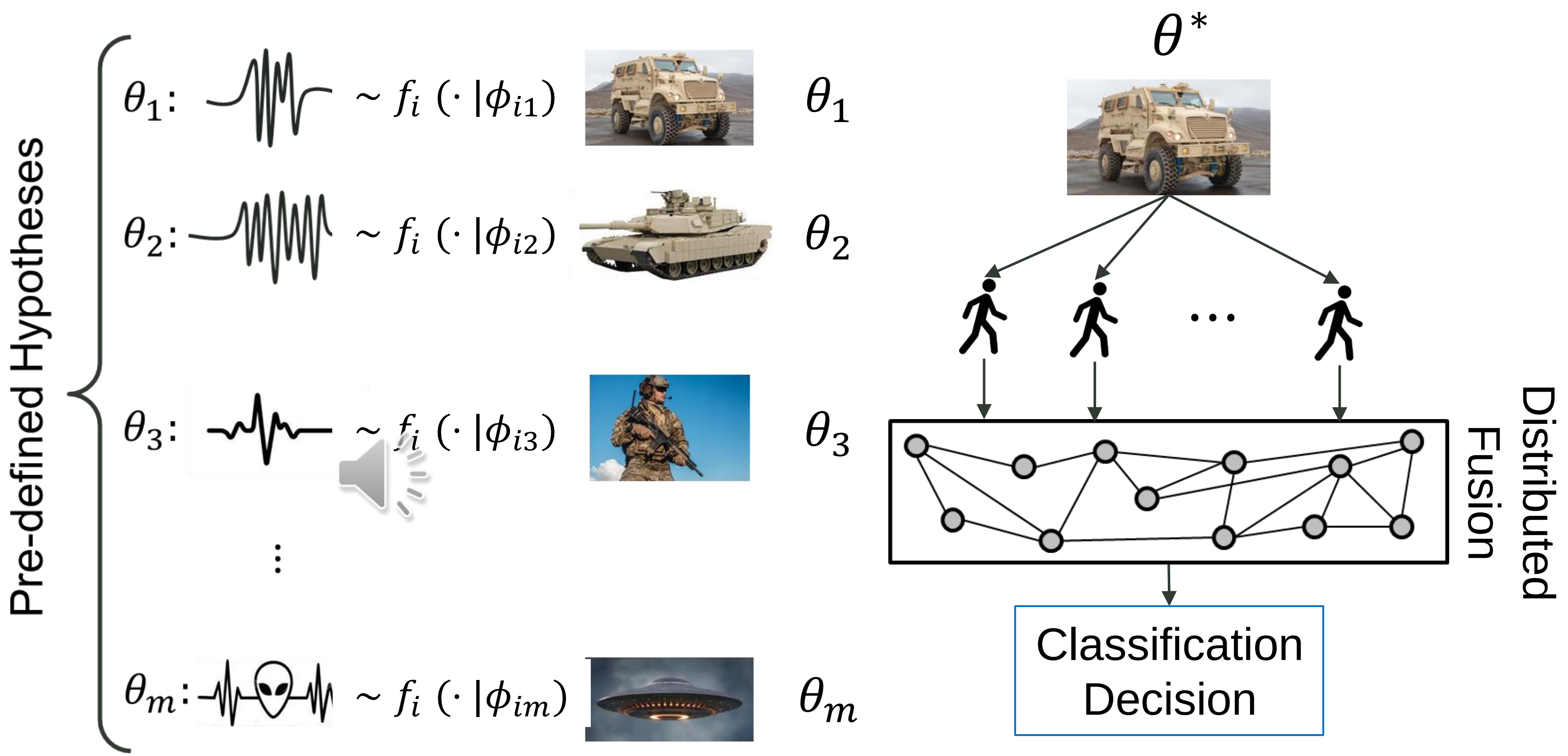




Updated Generative EDL Method

EPISTEMIC UNCERTAINTY IN HYPOTHESIS TESTING





Objective

Develop a distributed fusion framework that identifies the best hypothesis θ^* when the corresponding likelihood parameters ϕ_{ik} are uncertain

EXPRESSIONS OF BELIEF IN HYPOTHESES

Traditional Approaches

- The likelihood ratio test

$$\Lambda_\theta = \frac{P(X_{1:t}|\hat{\pi}_\theta)}{\sum_\theta P(X_{1:t}|\hat{\pi}_\theta)}$$

$$\theta^* = argmax_\theta \Lambda_\theta$$

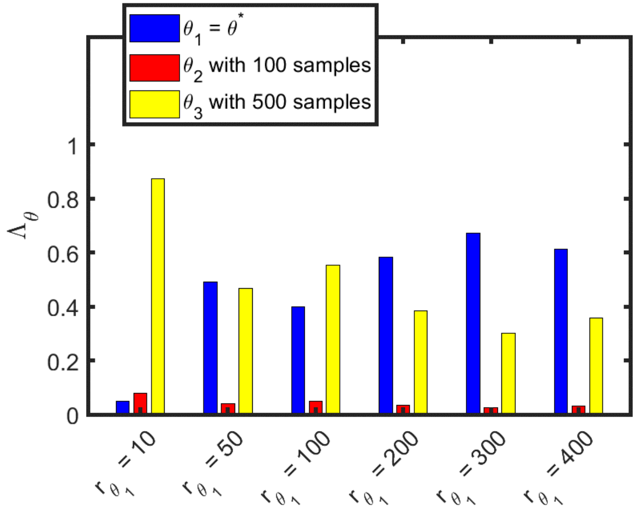
Uncertain Models

- The Uncertain Likelihood Ratio test (ULRT)

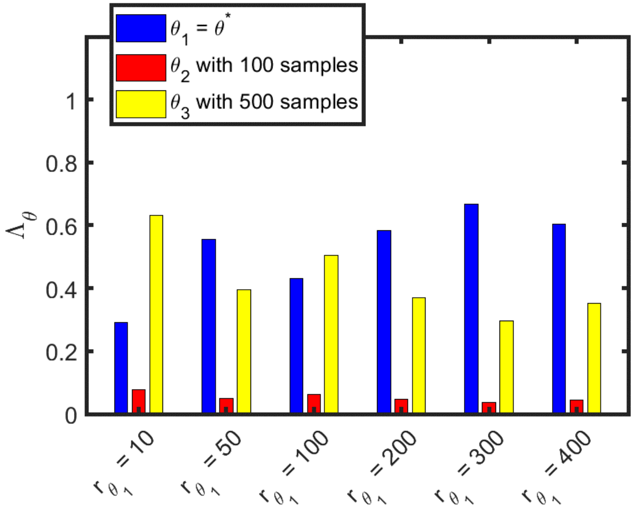
$$\Lambda_\theta = \frac{\int P(X_{1:t}|\pi) f(\pi|r_\theta) d\pi}{\int P(X_{1:t}|\pi) f_0(\pi) d\pi}$$

- Interpretation:
  - $\Lambda_\theta \gg 1$: Class θ is consistent w/ the ground truth
  - $\Lambda_\theta \ll 1$: Class θ is inconsistent w/ the ground truth
  - $\Lambda_\theta \approx 1$: Cannot make a determination

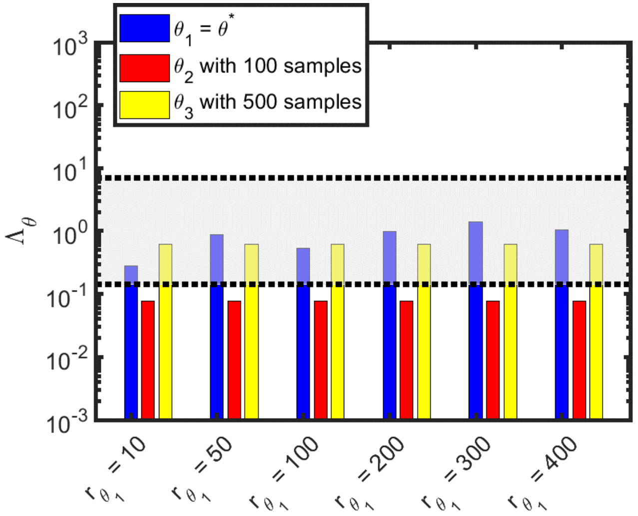
10 observations



Inference with Likelihood Ratio Test



ULR – Best Hypothesis



ULRT – Admissible Hypotheses

EXPRESSIONS OF BELIEF IN HYPOTHESES

Traditional Approaches

- The likelihood ratio test

$$\Lambda_\theta = \frac{P(X_{1:t}|\hat{\pi}_\theta)}{\sum_\theta P(X_{1:t}|\hat{\pi}_\theta)}$$

$$\theta^* = argmax_\theta \Lambda_\theta$$

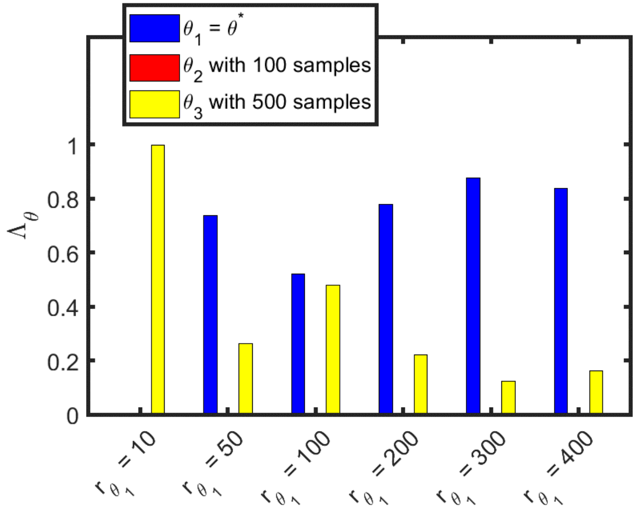
Uncertain Models

- The Uncertain Likelihood Ratio test (ULRT)

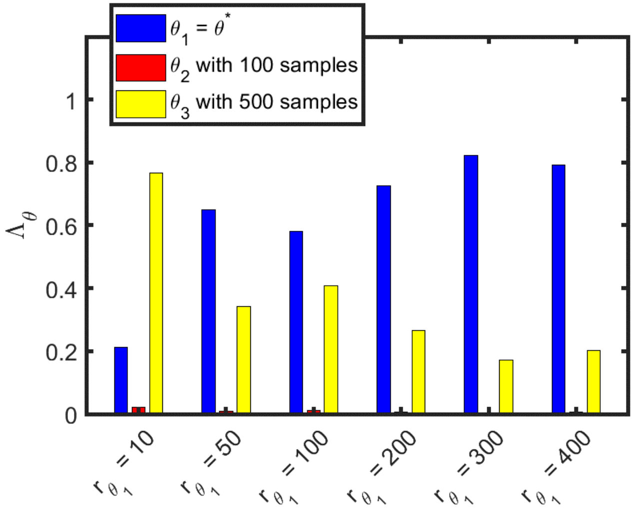
$$\Lambda_\theta = \frac{\int P(X_{1:t}|\pi) f(\pi|r_\theta) d\pi}{\int P(X_{1:t}|\pi) f_0(\pi) d\pi}$$

- Interpretation:
  - $\Lambda_\theta \gg 1$: Class θ is consistent w/ the ground truth
  - $\Lambda_\theta \ll 1$: Class θ is inconsistent w/ the ground truth
  - $\Lambda_\theta \approx 1$: Cannot make a determination

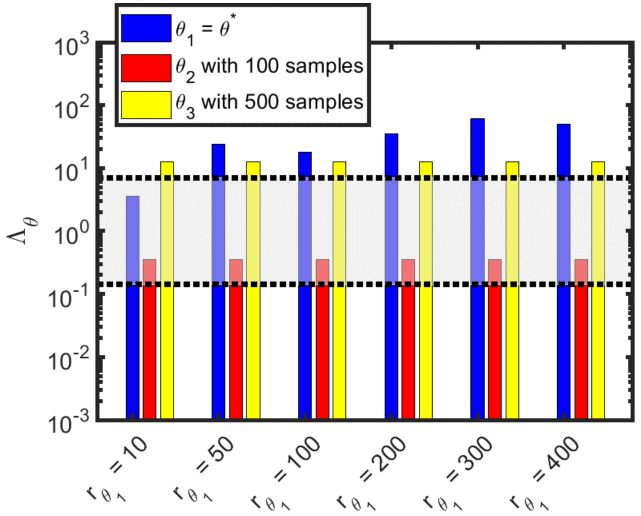
100 observations



Inference with Likelihood Ratio Test



ULR – Best Hypothesis



ULRT – Admissible Hypotheses

EXPRESSIONS OF BELIEF IN HYPOTHESES

Traditional Approaches

- The likelihood ratio test

$$\Lambda_\theta = \frac{P(X_{1:t}|\hat{\pi}_\theta)}{\sum_\theta P(X_{1:t}|\hat{\pi}_\theta)}$$

$$\theta^* = argmax_\theta \Lambda_\theta$$

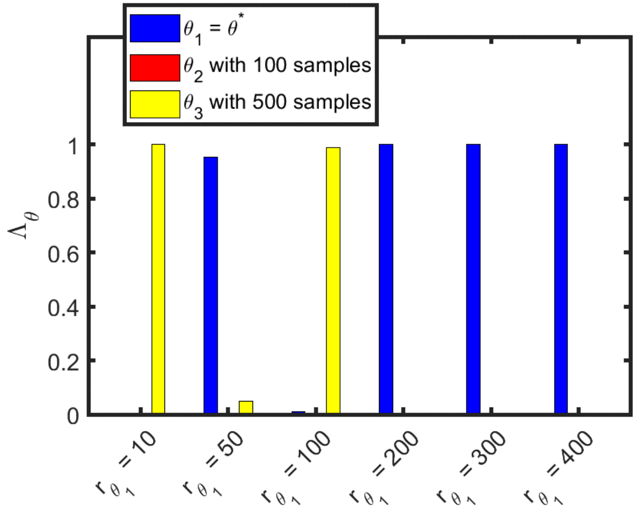
Uncertain Models

- The Uncertain Likelihood Ratio test (ULRT)

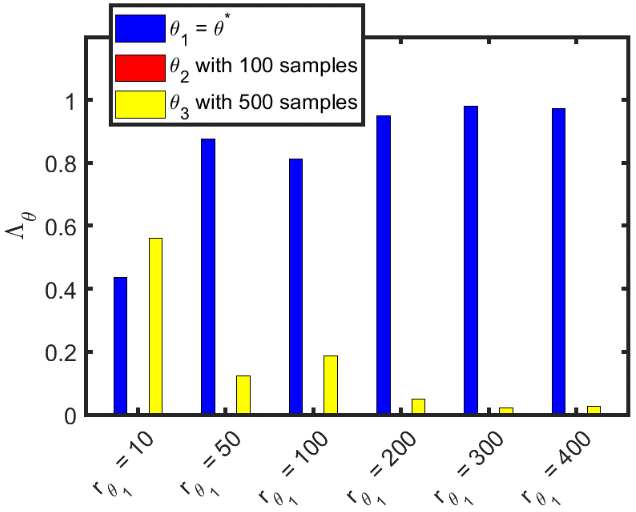
$$\Lambda_\theta = \frac{\int P(X_{1:t}|\pi) f(\pi|r_\theta) d\pi}{\int P(X_{1:t}|\pi) f_0(\pi) d\pi}$$

- Interpretation:
  - $\Lambda_\theta \gg 1$: Class θ is consistent w/ the ground truth
  - $\Lambda_\theta \ll 1$: Class θ is inconsistent w/ the ground truth
  - $\Lambda_\theta \approx 1$: Cannot make a determination

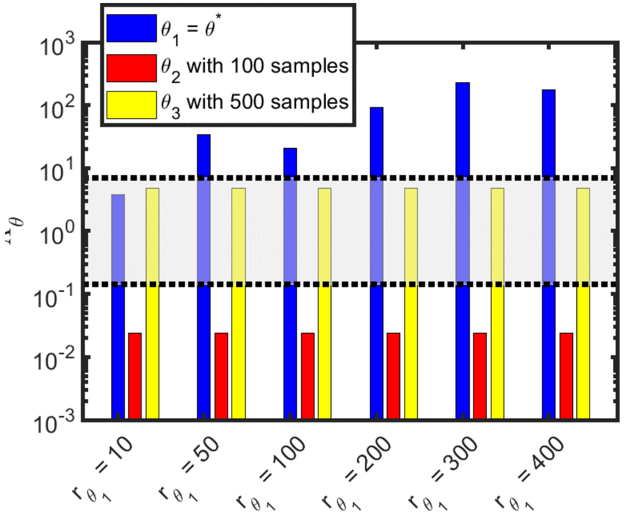
1000 observations



Inference with Likelihood Ratio Test

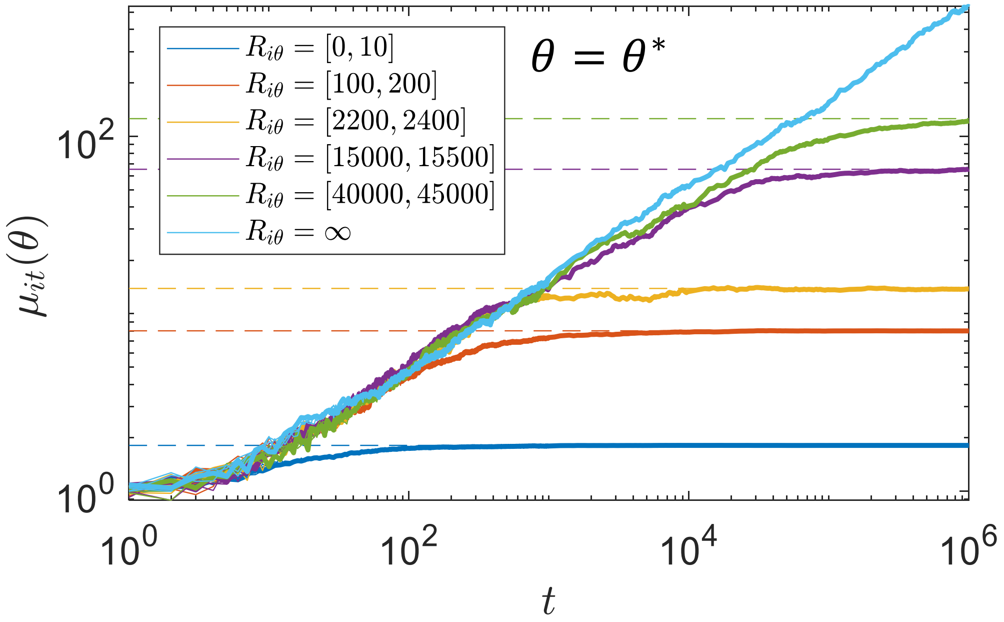


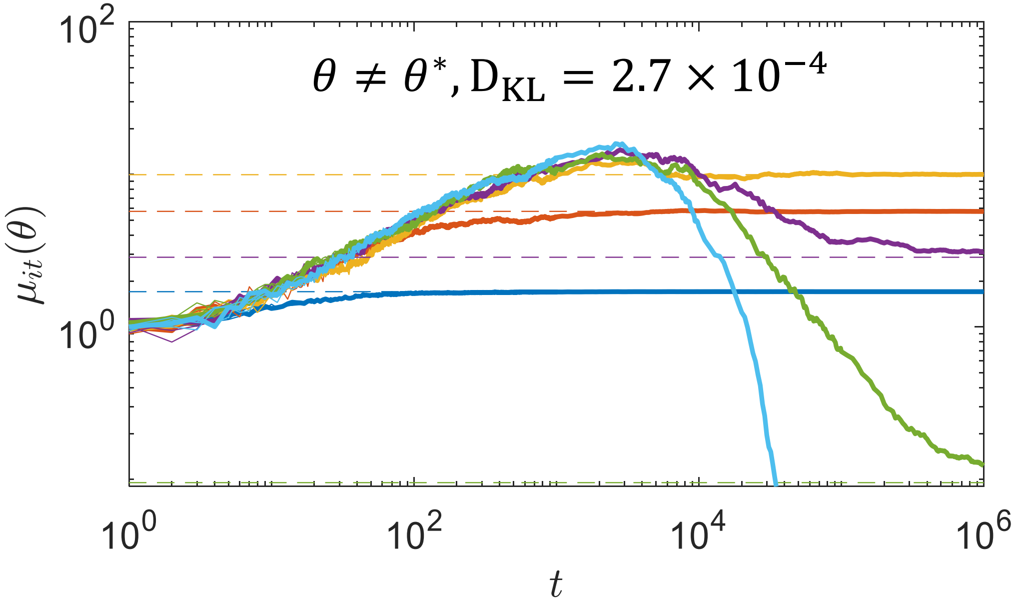
ULR – Best Hypothesis



ULRT – Admissible Hypotheses

GENERAL RESULTS WITH UNCERTAIN MODELS





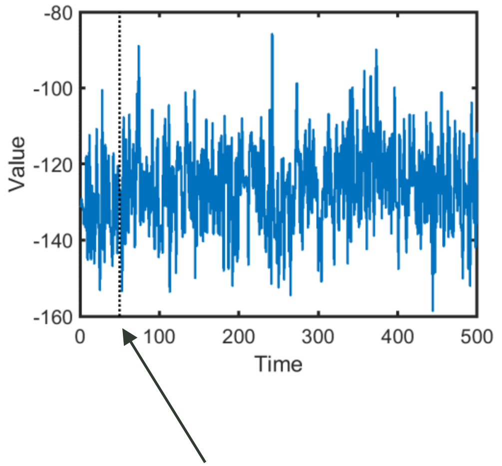
- ✓ Beliefs converge to:

$$\lim_{t\to\infty} \mu_{it}(\theta) = \left(\prod_{i=1}^{m} \widetilde{\Lambda}_{i\theta}\right)^{\frac{1}{m}}$$

- ✓ With a precise prior evidence (i.e., $R_{i\theta} = \infty$):
  - $\lim_{t\to\infty} \mu_{it}(\theta) = \infty$ when $\theta = \theta^*$ for all agents, and
  - $\lim_{t\to\infty} \mu_{it}(\theta) = 0$ when $\theta \neq \theta^*$ for at least one agent

- ✓ Results hold for:
  - Static Undirected Graphs w/ Doubly Stochastic A Matrix
  - B-connected Time-Varying Directed Graphs
  - Communication Constrained learning
  - Misspecified Distributions
  - Non-parametric distributions translated to Multinomial uncertain models via binning
  - Adversarial attacks
  - Active Learning

QUICKEST CHANGE DETECTION

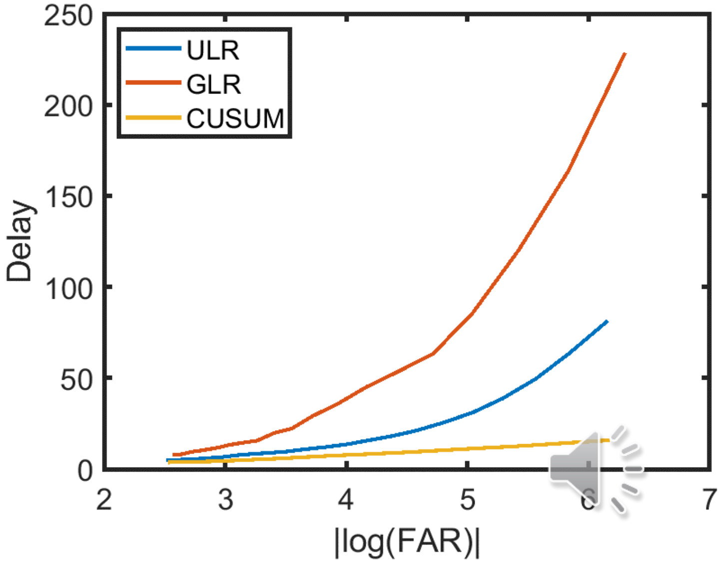


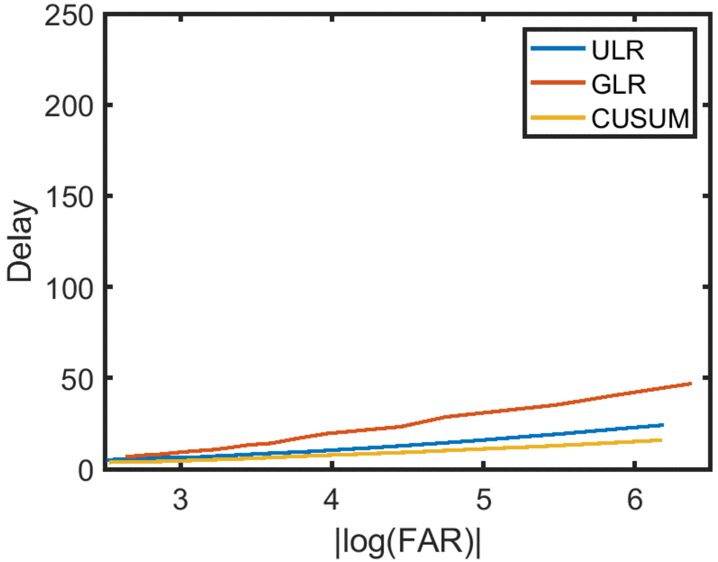
change point

$$D_{KL} = 0.1$$

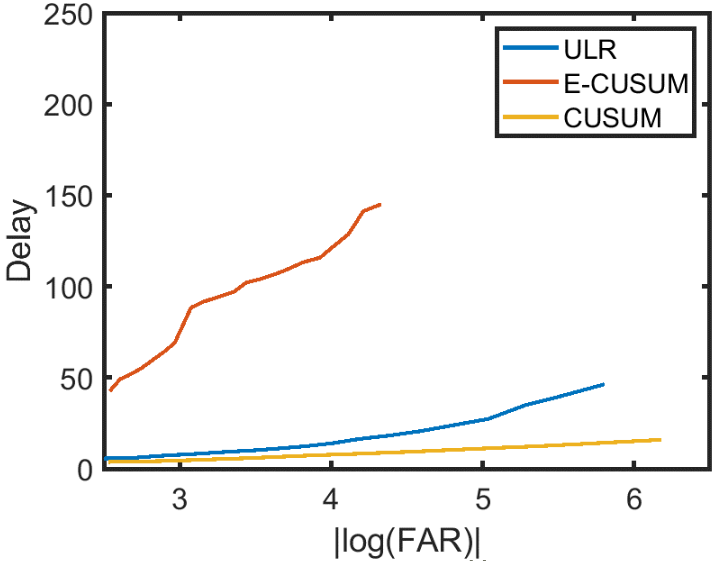
Uncertain Prechange | Known Prechange

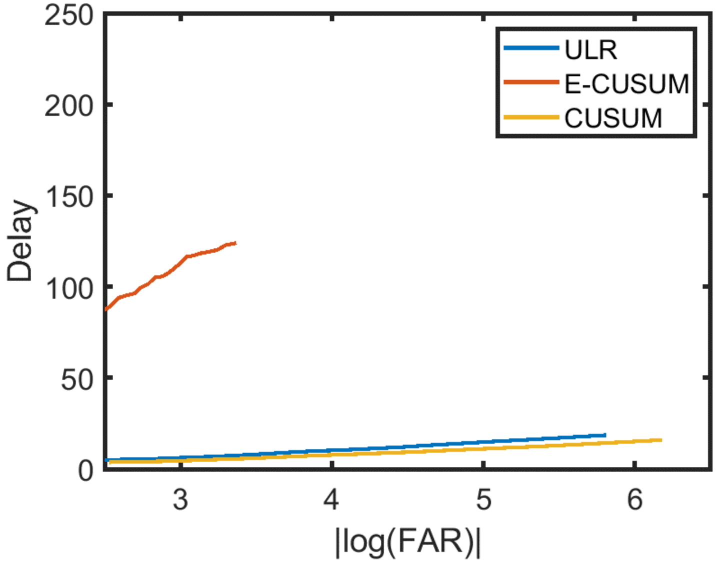
No Side-channel Evidence



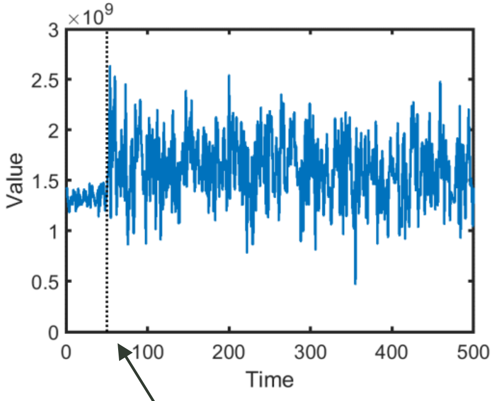


Evidence $|r_\theta| = 10$





QUICKEST CHANGE DETECTION



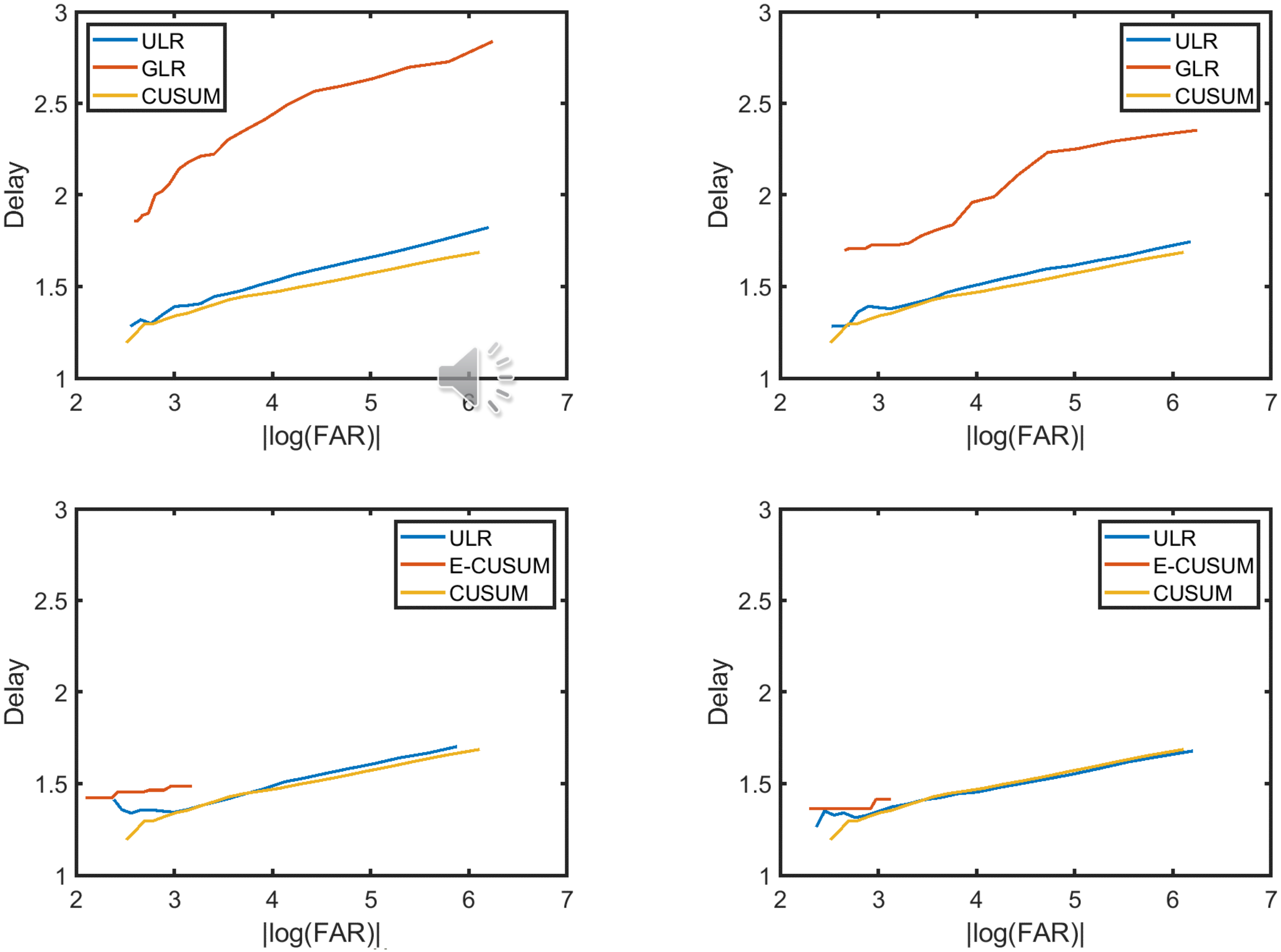
$$D_{KL} = 1.0$$

Uncertain Prechange

Known Prechange

No Side-channel Evidence

Evidence $|r_\theta| = 10$



CONCLUSION

Summary

- Learning: Extraction of epistemic uncertainty during learning
- Inference: Approximate methods to determine epistemic uncertainty of query answers
- Hypothesis Testing: Beliefs that capture epistemic uncertainty
- Tradeoff of higher computational complexity to capture epistemic uncertainty

Ways Forward

- Theory: Performance (e.g., DeCBoD) guarantees for the approximations
- Understand when epistemic uncertainty is required
- End to end computation of epistemic uncertainty through neuro-symbolic architectures
- Uncertain hypothesis testing over larger dimensional observations possibly using deep learning likelihood model
- Epistemic uncertainty for multiple target tracking

MAIN REFERENCES

- Lance Kaplan and Magdalena Ivanovska. "Efficient belief propagation in second-order Bayesian networks for singly-connected graphs." *International Journal of Approximate Reasoning,* vol. 93 pp. 132-152. 2018
- Federico Cerutti, Lance M. Kaplan, Angelika Kimmig, and Murat Sensoy. "Handling Epistemic and Aleatory Uncertainties in Probabilistic Circuits." *Machine Learning* 111, no. 4, pp. 1259-1301. 2022.
- Murat Sensoy, Lance Kaplan, Federico Cerutti, and Maryam Saleki. "Uncertainty-aware deep classifiers using generative models." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 5620-5627. 2020.
- Murat Sensoy, Lance Kaplan, and Melih Kandemir. "Evidential deep learning to quantify classification uncertainty." In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 3183-3193. 2018.
- James Z. Hare, Cesar A. Uribe, Lance Kaplan, and Ali Jadbabaie. "Non-Bayesian social learning with uncertain models." *IEEE Transactions on Signal Processing,* vol. 68, pp. 4178-4193. 2020.
- James Z. Hare, Cesar A. Uribe, Lance Kaplan, and Ali Jadbabaie. "A general framework for distributed inference with uncertain models." *IEEE Transactions on Signal and Information Processing over Networks*, vol. 7, pp. 392-405. 2021.

RECENT PAPERS ON UNCERTAIN BAYESIAN NETWORKS

- Conrad Hougen, Lance M. Kaplan, Federico Cerutti, and Alfred O. Hero. "Uncertain Bayesian Networks: Learning from Incomplete Data." In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing,* 2021.
- Conrad Hougen, Lance M. Kaplan, Magdalena Ivanovska, Federico Cerutti, Kumar Vijay Mishra, and Alfred O. Hero. "SOLBP: Second-Order Loopy Belief Propagation for Inference in Uncertain Bayesian Networks." In *2022 25th International Conference on Information Fusion,* 2022.

RECENT PAPERS ON UNCERTAIN LIKELIHOOD MODELS

- James Z. Hare, César A. Uribe, Lance M. Kaplan, and Ali Jadbabaie. "On malicious agents in non-Bayesian social learning with uncertain models." In *2019 22th International Conference on Information Fusion (FUSION)*. 2019.
- César A. Uribe, James Z. Hare, Lance Kaplan, and Ali Jadbabaie. "Non-Bayesian social learning with uncertain models over time-varying directed graphs." In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 3635-3640. 2019.
- James Z. Hare, César A. Uribe, Lance M. Kaplan, and Ali Jadbabaie. "Communication constrained learning with uncertain models." In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8609-8613. 2020.
- James Z. Hare, Lance Kaplan, and Venugopal V. Veeravalli. "Toward Uncertainty Aware Quickest Change Detection." In *2021 IEEE 24th International Conference on Information Fusion (FUSION)*. 2021.
- James Z. Hare, César A. Uribe, Lance Kaplan, and Ali Jadbabaie. "Toward Active Sequential Hypothesis Testing with Uncertain Models." In *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 3709-3716. 2021.