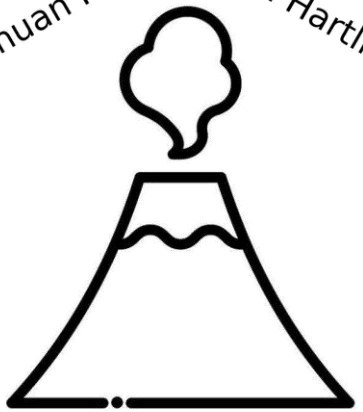
Northwestern University

George Mason University Department of Computer Science

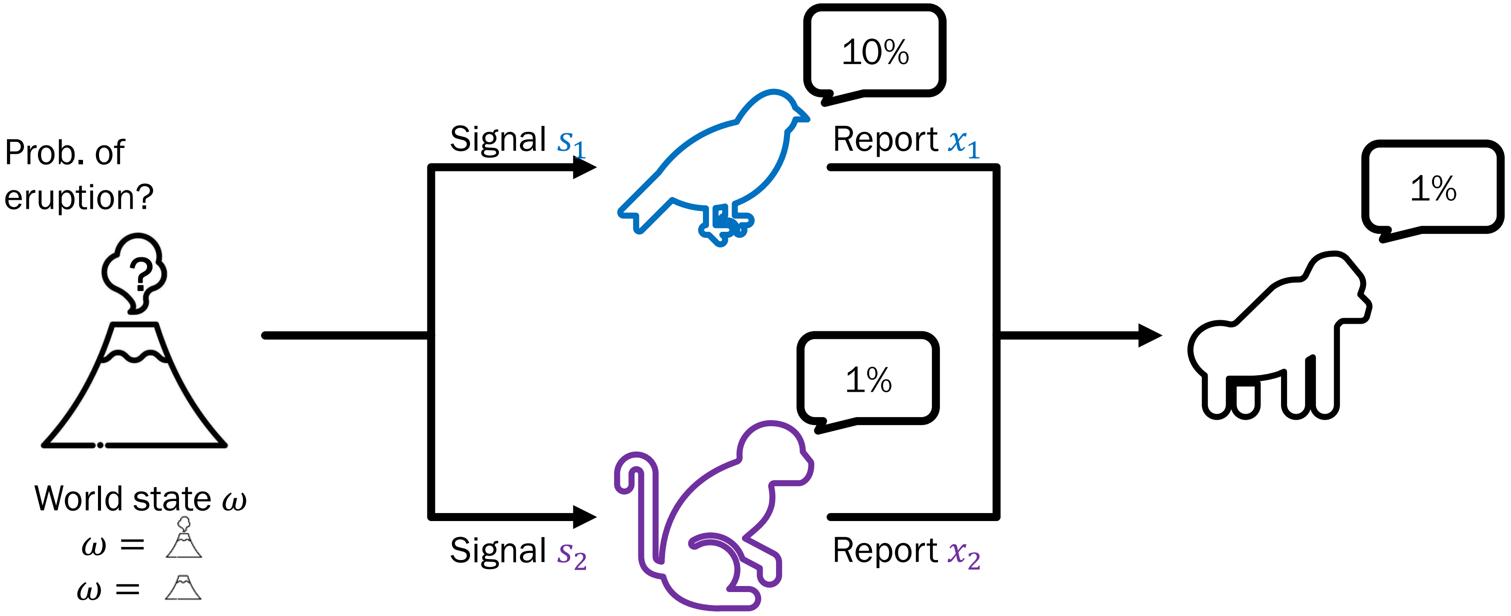
Algorithmic Robust Forecast Aggregation



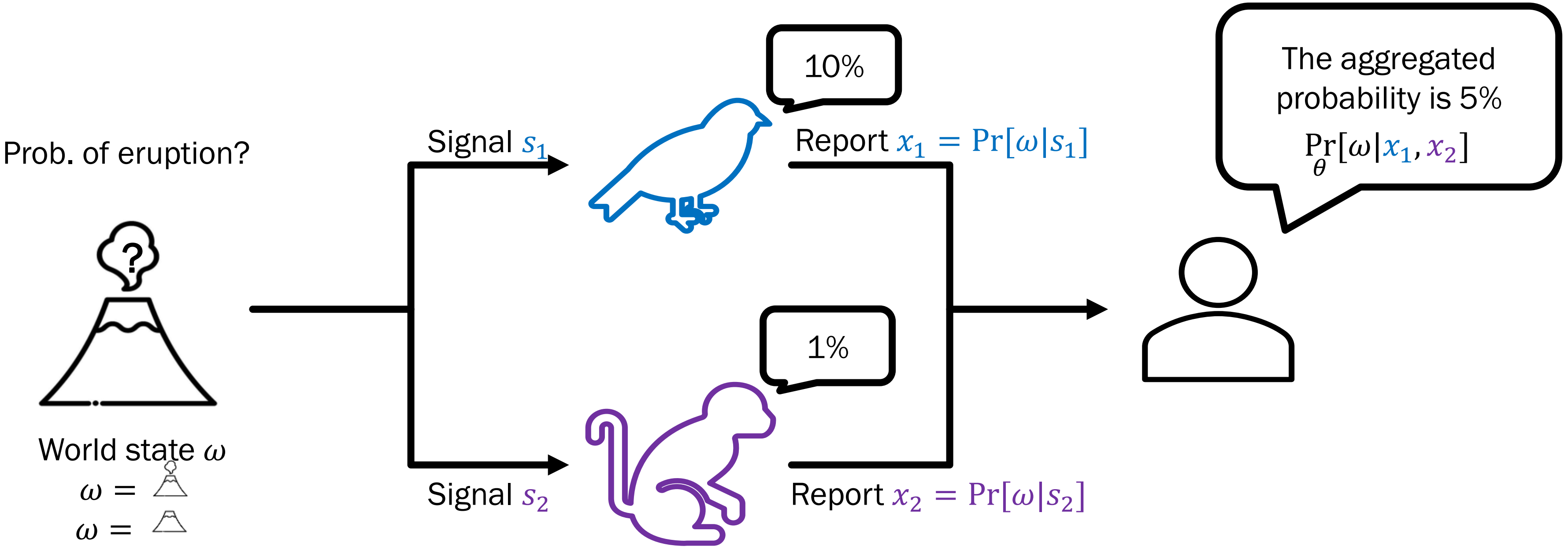
Zhihuan Huang, Jason Hartline, Yongkang Guo, FangYi-Yu, Anant Shah, Yuqing Kong

北京大学前沿计算研究中心
Center on Frontiers of Computing Studies, Peking University

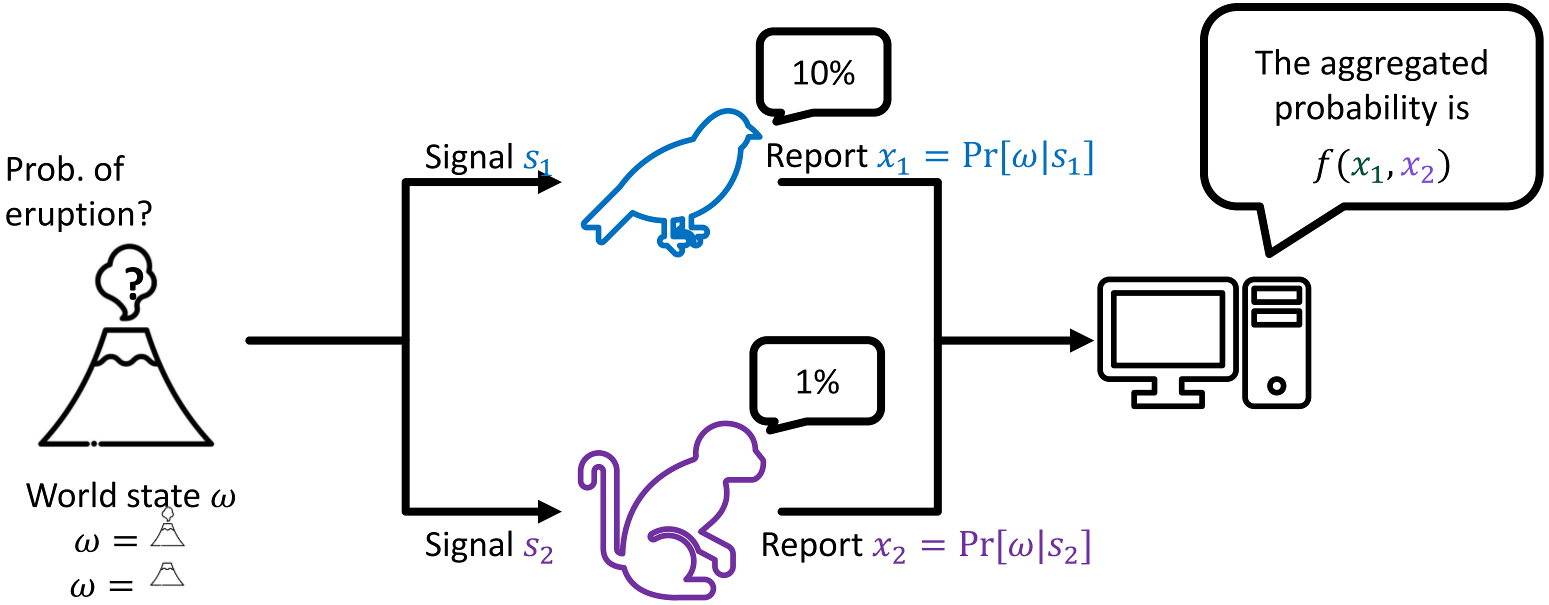
Long time ago…



When human master the Bayes rule…

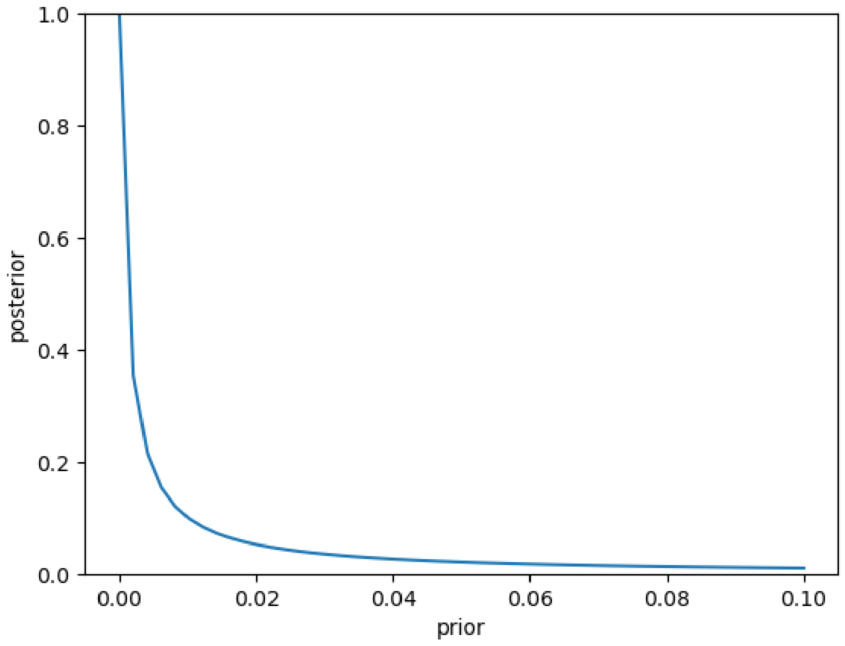


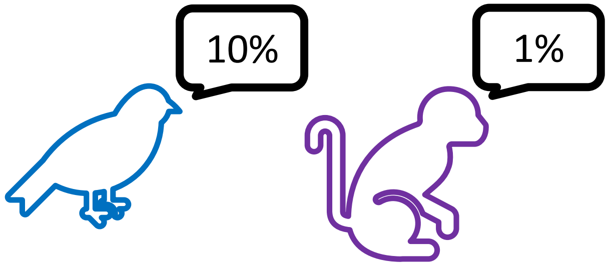
When we do not know joint distribution θ...



Prior matters

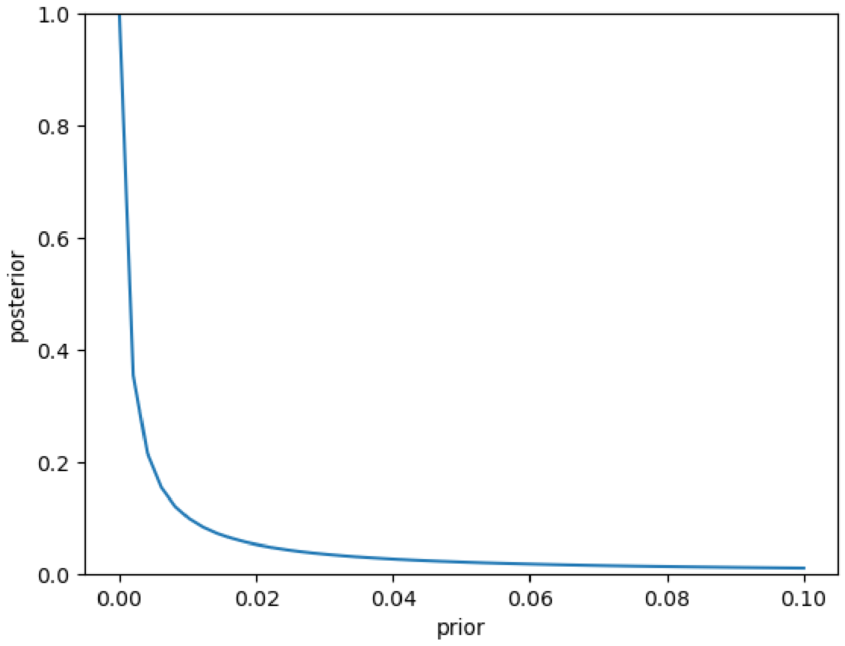
- If the prior probability of eruption μ
  - Given prediction x_1 and x_2, the posterior is

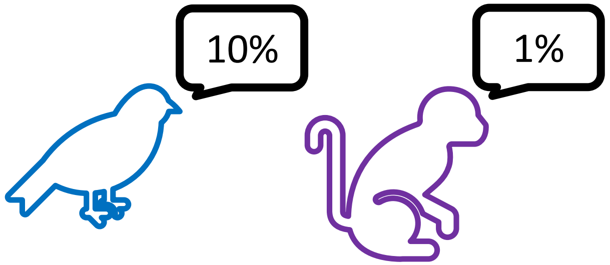




Prior Matters

- If the prior probability of eruption μ
  - Given prediction x_1 and x_2, the posterior is
$$\frac{(1-\mu)x_1x_2}{(1-\mu)x_1x_2+\mu(1-x_1)(1-x_2)}$$
  - 5% => safe
  - 0.000001% => **far less than** the forecasts => Dangerous!





Aggregating multiple forecasts

- Weather forecast
- Elections
- Investments
- Medical Prognosis

Outline

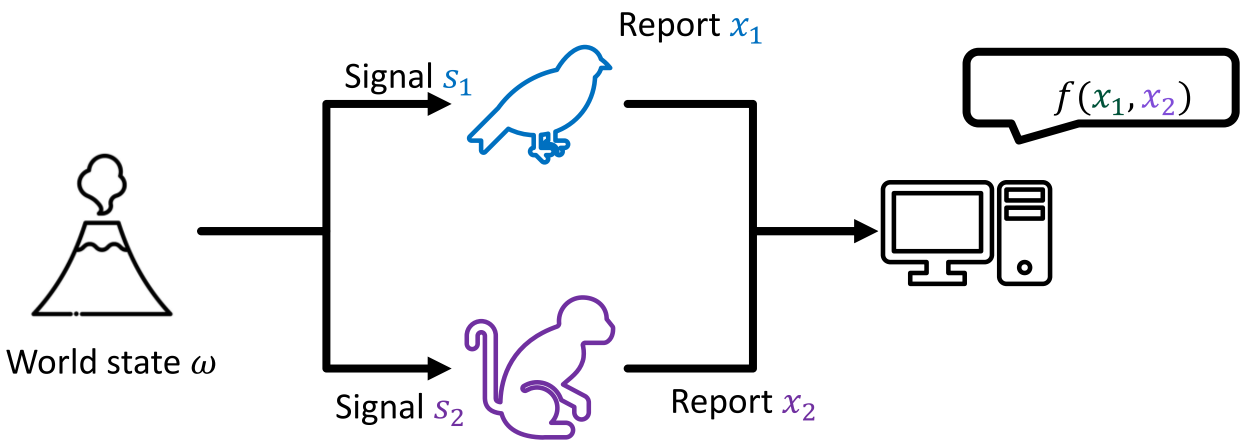
- Robust forecast aggregation problem
  - An ignorant aggregator
    - does not know the information structure
    - aims to minimize the regret in the worst information structures.
- Theoretical results
  - FPTAS based on online learning method
  - Two challenges
- Numerical results

Outline

- Robust forecast aggregation problem
  - An ignorant aggregator
    - does not know the information structure
    - aims to minimize the regret in the worst information structures.
- Theoretical results
  - FPTAS based on online learning method
  - Two challenges
- Numerical results

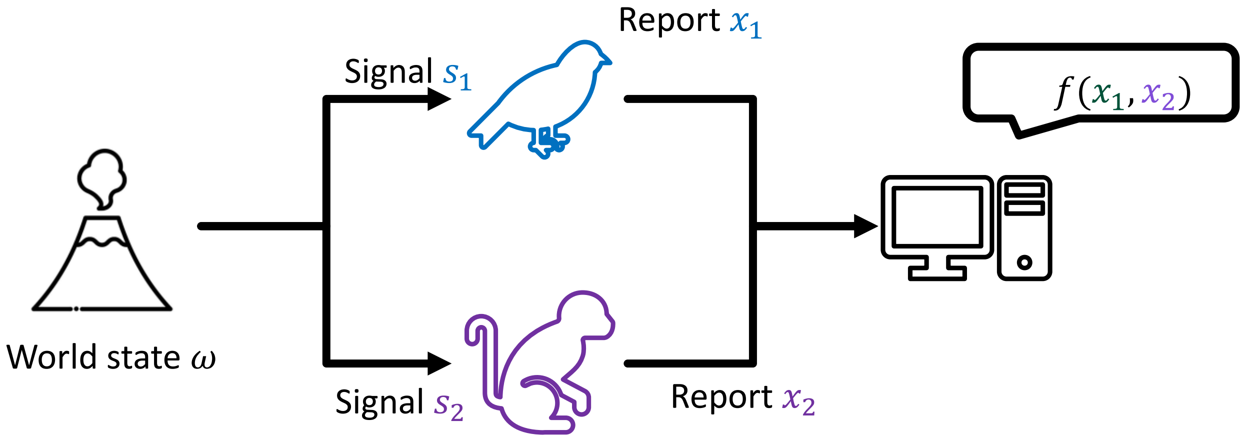
Robust Forecast Aggregation Problem [ABS18]

- Inputs
  - Binary state of nature $\Omega = \{0,1\}$
  - Two forecasters $\{1,2\}$ with signal sets S_1, S_2.
  - Information structure
    $$\theta \in \Delta(\Omega \times S_1 \times S_2)$$
  - Each forecaster reports the posterior
    $$x_i = \Pr_\theta[\omega = 1 | s_i]$$



Robust Forecast Aggregation Problem

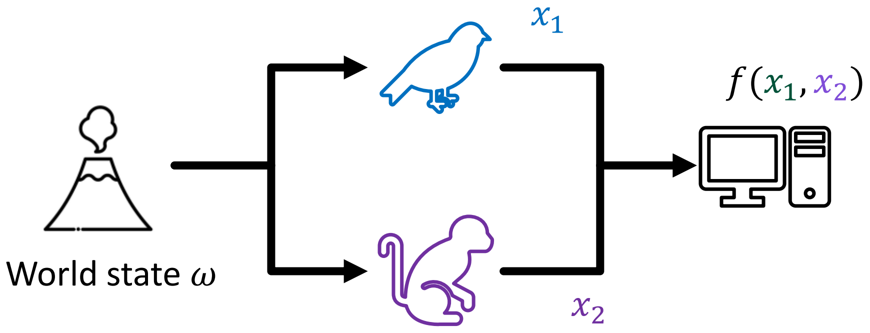
- Inputs
  - Binary state of nature $\Omega = \{0,1\}$
  - Two forecasters $\{1,2\}$ with signal sets S_1, S_2.
  - Information structure
    $$\theta \in \Delta(\Omega \times S_1 \times S_2)$$
  - Each forecaster reports the posterior
    $$x_i = \Pr_\theta[\omega = 1 | s_i]$$
- Output a forecast $f(x_1, x_2)$ where $f: [0,1]^2 \to [0,1]$ is called an aggregation scheme



Measure aggregation schemes

The forecast should be accurate and robust

- Accurate:
  - Squared loss function $l(f,\theta) = \mathrm{E}_\theta[(f - w)^2]$,
  - Regret against Bayesian aggregator with θ
    $R(f,\theta) = l(f,\theta) - l(opt_\theta,\theta)$.

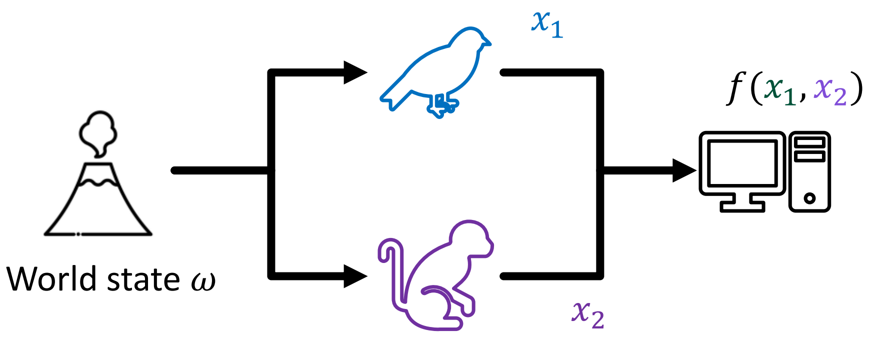


Measure aggregation schemes

The forecast should be accurate and robust

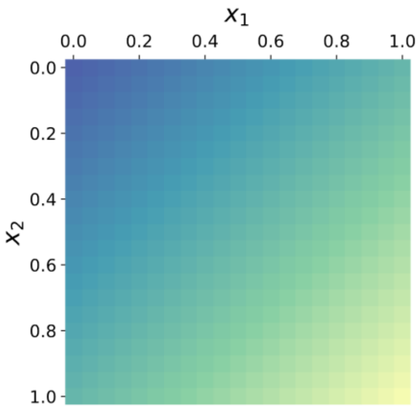
- Accurate:
  - Squared loss function $l(f,\theta) = \mathrm{E}_\theta[(f-w)^2]$,
  - Regret against Bayesian aggregator with θ
    $R(f,\theta) = l(f,\theta) - l(opt_\theta,\theta)$.
- Robust: a set of information structures Θ the worst-case performance is small

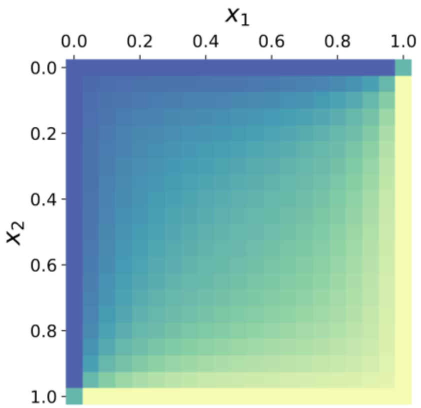
$$\inf_{f\in\mathcal{F}} \sup_{\theta\in\Theta} R(f,\theta)$$

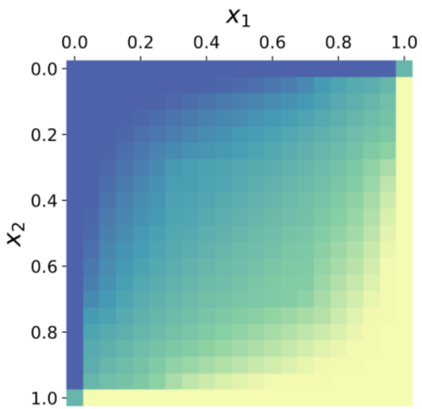


Possible Approaches

- Common function, e.g., average
- Delicate designed aggregators [ABS, PNAS 2018]
- An algorithmic and systematic method?







Outline

- Robust forecast aggregation problem
  - An ignorant aggregator
    - does not know the information structure
    - aims to minimize the regret in the worst information structures.
- Theoretical results
  - FPTAS based on online learning method
  - Two challenges
- Numerical results

FPTAS for optimal forecasts aggregation

There is an efficient algorithm outputting an ϵ-optimal aggregator for $\inf_{f \in \mathcal{F}} \sup_{\theta \in \Theta} R(f, \theta)$ if one of the following holds

1. Θ is finite
2. Θ consists of conditional independent information structures and $\mathcal{F}$ is the collection of L-Lipschitz aggregators

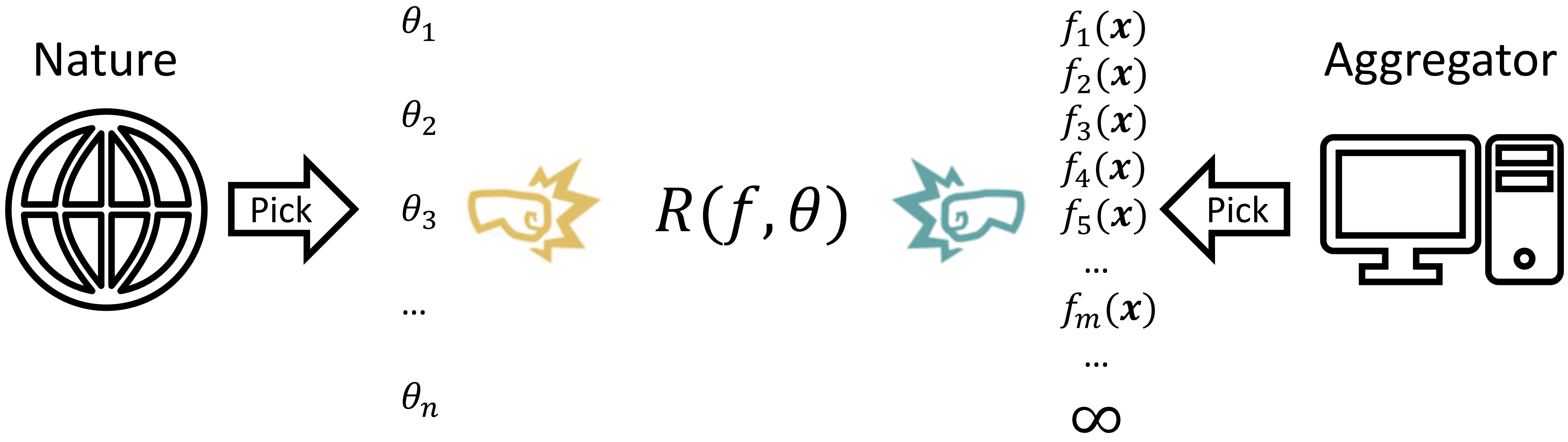
FPTAS for optimal forecasts aggregation

There is an efficient algorithm outputting an ϵ-optimal aggregator for $\inf_{f \in \mathcal{F}} \sup_{\theta \in \Theta} R(f, \theta)$ if one of the following holds

1. Θ is finite
2. Θ consists of conditional independent information structures and $\mathcal{F}$ is the collection of L-Lipschitz aggregators

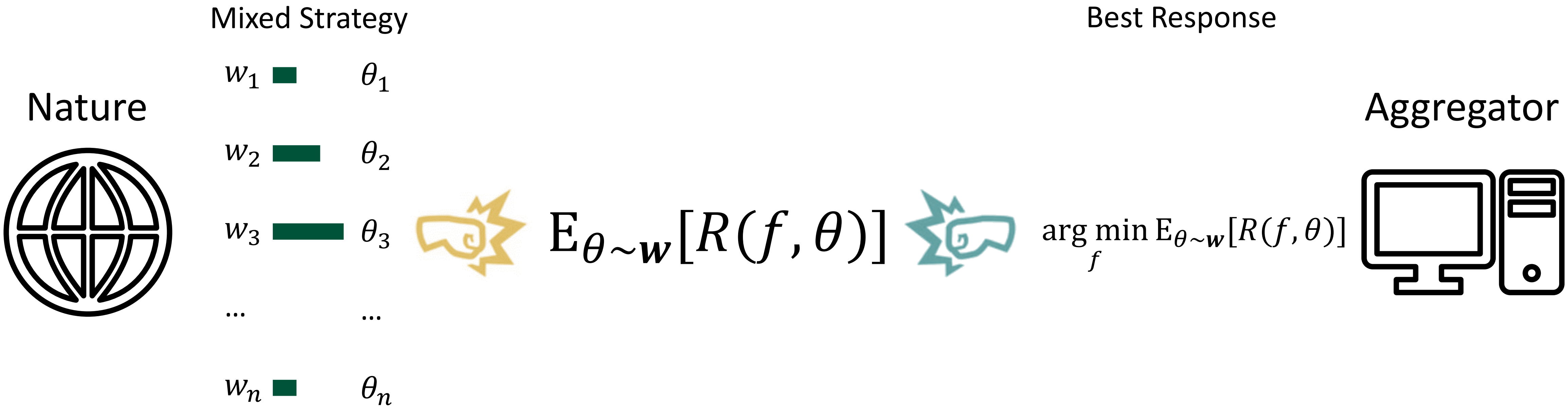
Algorithm: equilibrium computation

$\inf_{f \in \mathcal{F}} \sup_{\theta \in \Theta} R(f, \theta)$ as a zero-sum game



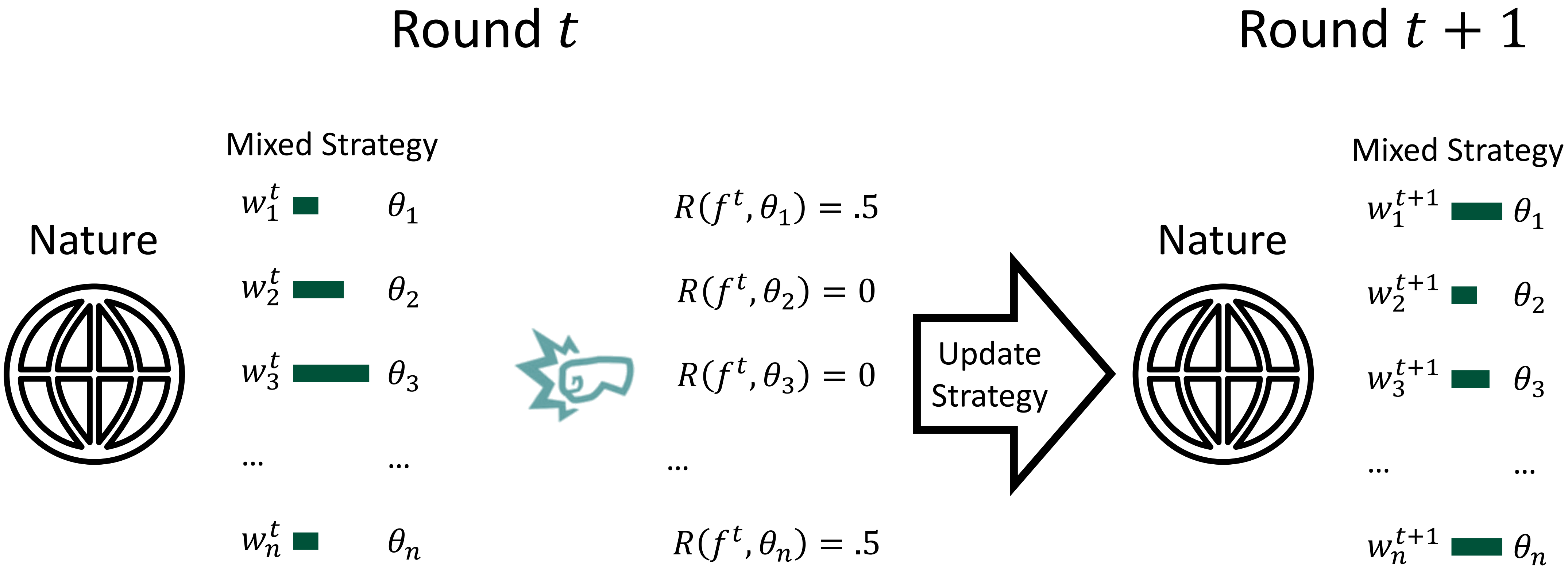
Online Learning: Efficient Best Response

No-regret vs Best response



Online Learning: Learn Mixed Strategy

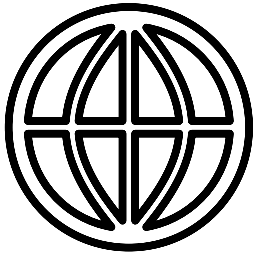
No-regret vs Best response



Online Learning: Optimal Solution

Nature

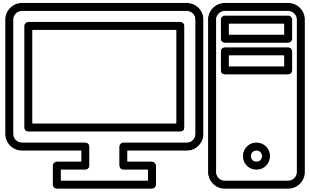
Aggregator



$$w^* = \frac{\sum_{t=1}^{T} w^t}{T}$$

$$\mathrm{E}_{\theta \sim w^*}[R(f^*, \theta)]$$

$$f^* = \frac{\sum_{t=1}^{T} f^t}{T}$$



$$\approx \inf_{f \in \mathcal{F}} \sup_{\theta \in \Theta} R(f, \theta)$$

FPTAS for optimal forecasts aggregation

There is an efficient algorithm outputting an ϵ-optimal aggregator for $\inf_{f \in \mathcal{F}} \sup_{\theta \in \Theta} R(f, \theta)$ if one of the following holds

1. Θ is finite
2. Θ consists of conditional independent information structures and $\mathcal{F}$ is the collection of L-Lipschitz aggregators

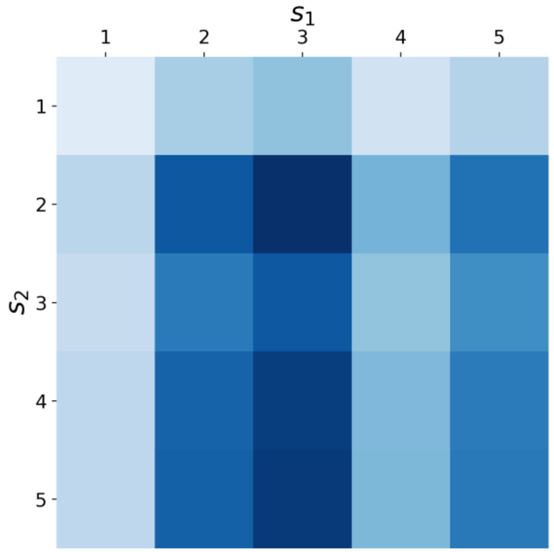
Two challenges

- The dimension of $\Theta \subset \Delta(\Omega \times S_1 \times S_2)$ is high.

Two challenges

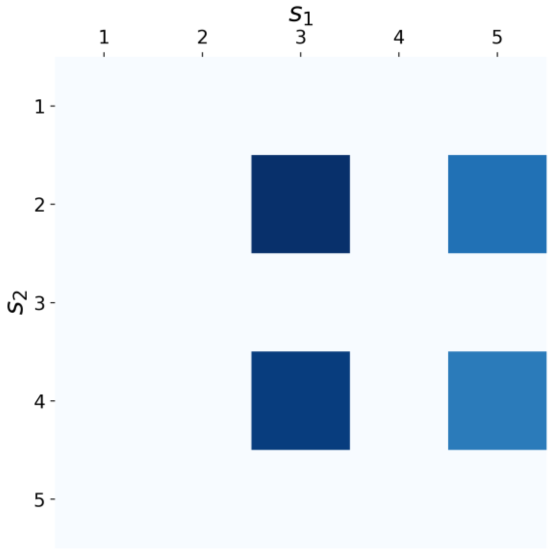
- The dimension of $\Theta \subset \Delta(\Omega \times S_1 \times S_2)$ is high.
  - Dimension reduction

Multiple signals



= Linear Combination of

Binary signals



s

Two challenges

- The dimension of $\Theta \subset \Delta(\Omega \times S_1 \times S_2)$ is high.
  - Dimension reduction
- Θ and $\mathcal{F}$ are continuous.
  - Discretize Θ to finite information structure Θ^{fin} and show $R(\Theta, \mathcal{F}) \approx R\left(\Theta^{fin}, \mathcal{F}\right)$

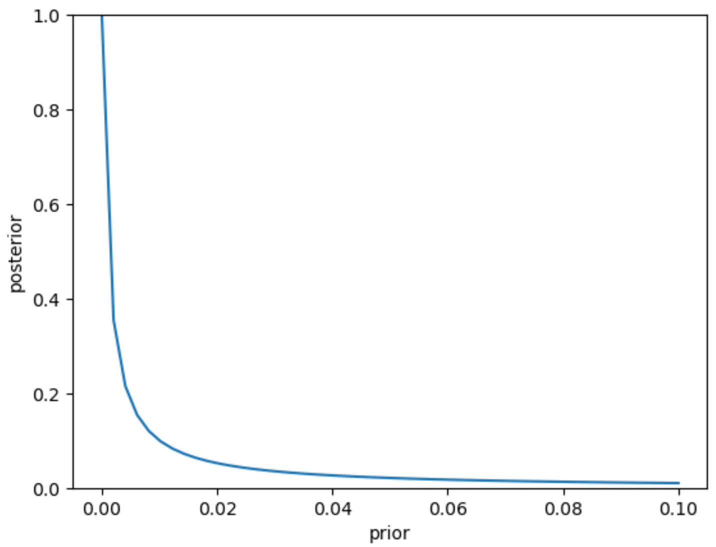
Control the Regret $R(\Theta, \mathcal{F}) \approx R\left(\Theta^{fin}, \mathcal{F}\right)$

- (ϵ, d)-Covering
  - For any $\theta \in \Theta$, we can find $\theta' \in \Theta'$ such that $d(\theta, \theta') \leq \epsilon$.
  - Use the nearest information structure in the Euclidean space.
  - Prove they are close in the TVD(EMD) metric.
- Discretize in two steps:
  - Discrete reports (EMD)
  - Discrete prior (TVD)

Control the Regret $R(\Theta, \mathcal{F}) \approx R\left(\Theta^{fin}, \mathcal{F}\right)$

$$R(f^t, \theta) = l(f^t, \theta) - l(\mathrm{opt}_\theta, \theta)$$

- Bayesian aggregator opt_θ is not smooth
  - – Find a smooth one to approximate opt_θ
  - – Trimming the non-smooth area.
  - – Extend the smooth regret to the full space.



- Best-response f^t is not smooth
  - – Restrict f^t to smooth function.
  - – Calculate the best smooth response by Ellipsoid method.

Two solutions

- Reduce dimension
  - The dimension of $\Theta \subset \Delta(\Omega \times S_1 \times S_2)$ is high.
- Discretize Θ and run no regret+best response
  - Calculate the best response f in $\mathcal{F}$.
  - Control the regret $R(\Theta, \mathcal{F}) \approx R\left(\Theta^{fin}, \mathcal{F}\right)$

Outline

- Robust forecast aggregation problem
  - An ignorant aggregator
    - does not know the information structure
    - aims to minimize the regret in the worst information structures.
- Theoretical results
  - FPTAS based on online learning method
  - Two challenges
- Numerical results

Our Results (Numerical)

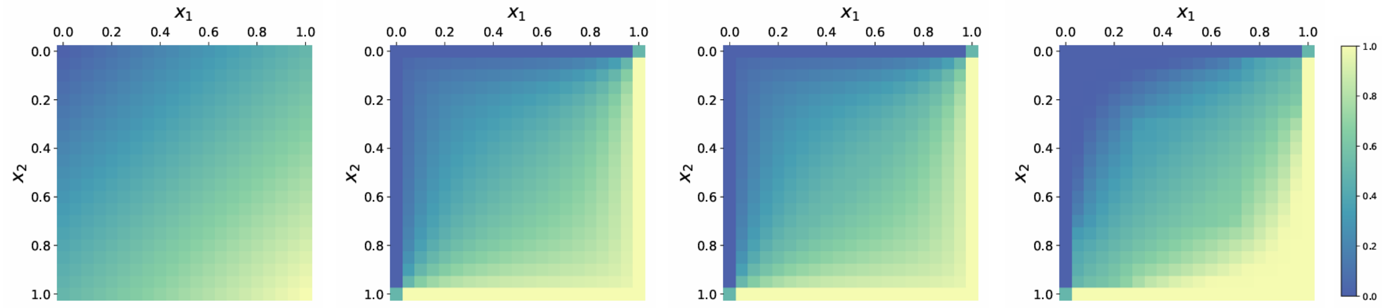
Aggregator	Formula	Regret
Simple average	$\frac{x_1+x_2}{2}$	0.0625
Average prior	$\frac{x_1x_2(1-\frac{x_1+x_2}{2})}{x_1x_2(1-\frac{x_1+x_2}{2})+(1-x_1)(1-x_2)\frac{x_1+x_2}{2}}$	0.0260
State-of-the-art	$\frac{x_1x_2(1-ep(x_1,x_2))}{x_1x_2(1-ep(x_1,x_2))+(1-x_1)(1-x_2)ep(x_1,x_2)}$	0.0250
Our	-	**0.0227**

Table 1: **Regret of different aggregators.**

Here $ep(x_1, x_2) = \begin{cases} 0.49x_1 + 0.49x_2, & if\ x_1 + x_2 \leq 1 \\ 0.49x_1 + 0.49x_2 + 0.02, & otherwise \end{cases}$

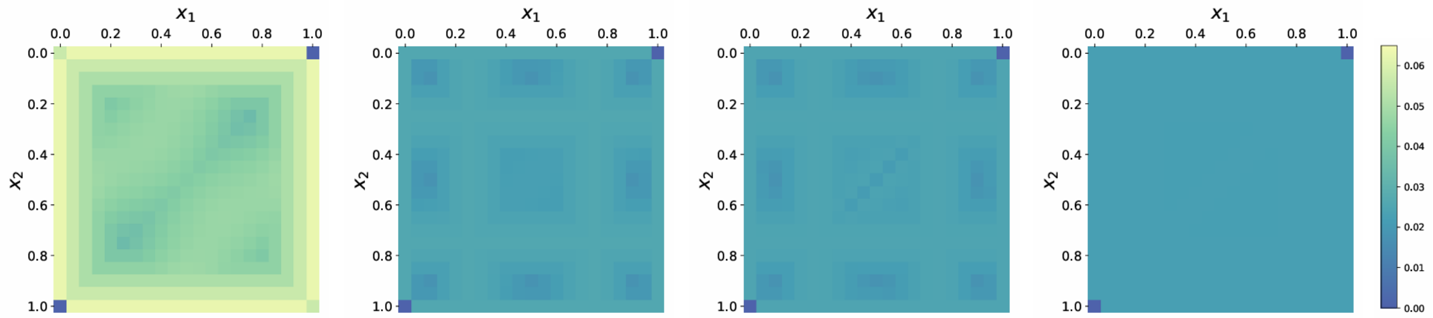
Our Results (Numerical)

- Extremization



(a) **Simple averaging** (b) **Average prior** (c) **State-of-the-art** (d) **Our aggregator**

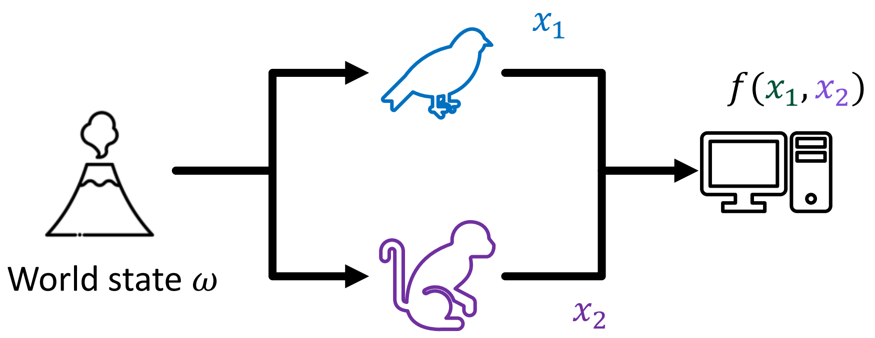
- Uniform loss



(a) **Simple averaging** (b) **Average prior** (c) **State-of-the-art** (d) **Our aggregator**

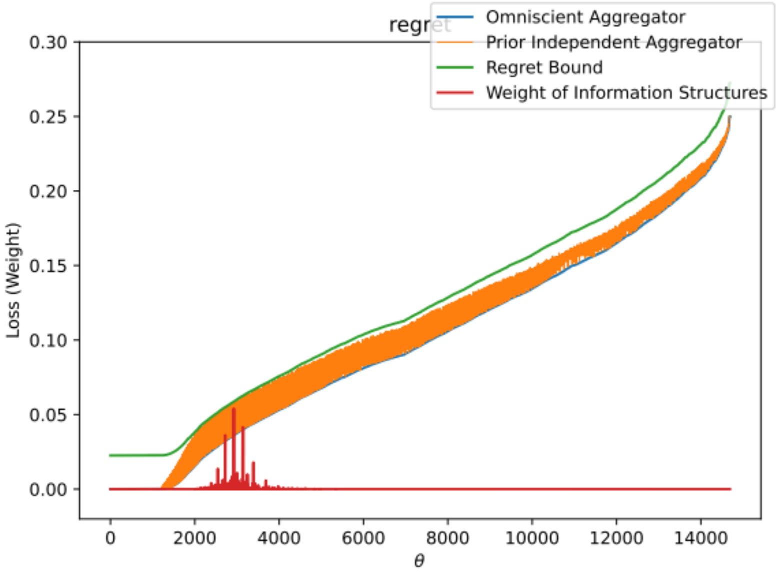
Conclusion and Future Work

- Conclusion
  - An algorithmic framework for robust information aggregation
  - Automatic design of aggregators, even in scenarios with complex report formats
- Future work using partial knowledge Θ
  - Aggregation
    - Higher order reports
    - Decision
  - Elicitation
    - Proper scoring rule

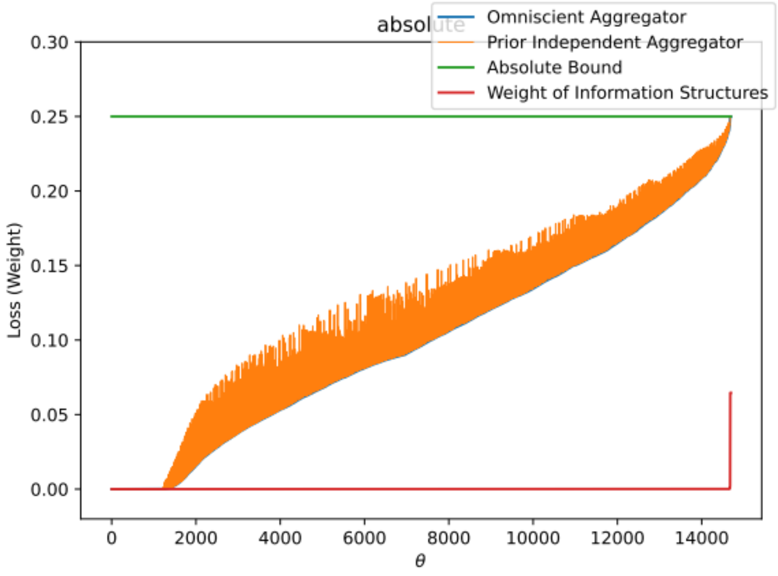


Additive vs. Ratio vs. Absolute

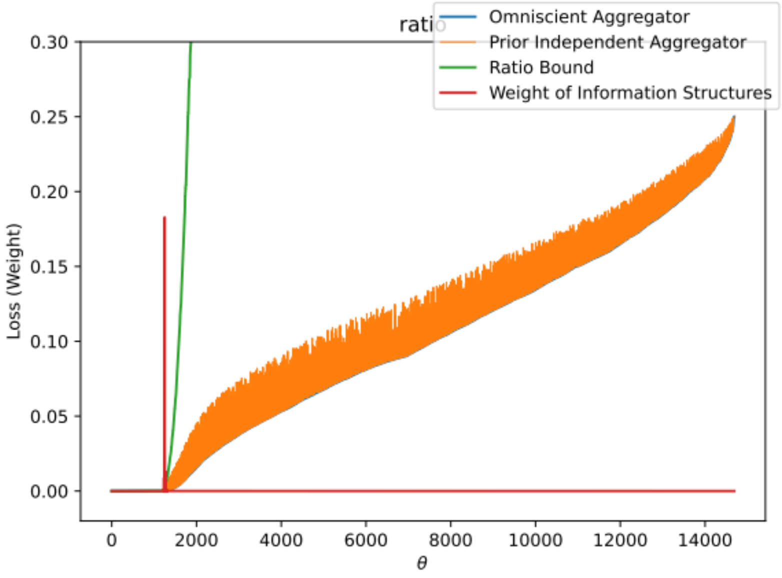
- $l(f^t, \omega) - l(\text{opt}_\theta, \omega)$



(a) **Additive**



(b) **Absolute**



(c) **Ratio**