Distributed Inference in Sensor Networks:

Part 1: Aggregating Probability Forecasts

Sanjeev Kulkarni

Department of Electrical Engineering

Princeton University

Joint work with Joel Predd, Dan Osherson, Vince Poor

NSF Workshop: Future Directions in Systems Research for Networked Sensing

Boston University, May 25-26, 2006

Distributed Inference in Sensor Networks:
Part 1: Aggregating Probability Forecasts

- Motivation, Background, Formulation
- Orthogonal Projection (Coherent Approximation Principle)
- Alternating Projection Algorithm
- Experiments

Part 2: Collaborative Regression (tomorrow, H.V. Poor)

Risk analysts & decision makers often confront an aggregation problem

"Las Vegas **probable** target"
"Biological attack **unlikely**"

-Judge 1

"Biological attack in Las Vegas

inconceivable"
"Target **most likely** Las Vegas"
-Judge 2

"Biological attack **virtually inevitable**"
"Las Vegas **anticipated** target for bio-terror."

-Judge 3

Are there scalable algorithms for fusing human forecasts of probability?

Wireless sensor networks for inference

Are there bandwidth & energy-efficient algorithms for distributed learning?

As computation/inference gets pushed to sensors, can we find good methods for hard/soft fusion (e.g., aggregating probability forecasts)?

Similar tools are useful for both problems

- Analogous structure and goals:

  local structure exploited for

  global computationally efficient aggregation

  communication-efficient learning in WSNs

- View solution as orthogonal projection

- Alternating projections provides successful approach to both problems

A Possible Methodology for Hard/Soft Fusion?

"Las Vegas **probable** target"
"Biological attack **unlikely**"

-Judge 1

"Biological attack in Las Vegas **inconceivable**"
"Target **most likely** Las Vegas"

-Judge 2

"Biological attack **virtually inevitable**"
"Las Vegas **anticipated** target for bio-terror."

-Judge 3

- Soft info can be valuable in conjunction with hard info.
  - What forecasts? Which sensors? How to ask/compute?
- Forecast/elicitation methodology paramount
  - Are expert/sensor forecasts meaningful? Calibrated? Biased?

Assumption:

Methodological issues have been properly addressed;

Probability forecasts require aggregation.

Notation

- ***Basic Events***

  given a vector of boolean variables $(X_1, X_2, \ldots, X_n)$

  e.g., X_1 = **[**Las Vegas Attacked**]**, X_2 = **[**Biological Attack**]**, ...

- ***Complex Events***

  Basic Events + $\{\neg, \vee, \wedge, \ldots\}$

  e.g., $\phi_1 = X_1 \wedge X_2$, $\phi_2 = \neg X_1 \vee X_3$,...

- ***Forecast***

  (Event, $\hat{p}$ *) with* $\hat{p} \in [0, 1]$

A set $\{(\phi_i, \hat{p}_i)\}_{i=1}^{m}$ of forecasts is **probabilistically coherent** iff they are implied by a joint distribution of probability over

$$(X_1, X_2, \ldots, X_n)$$

Incoherence and Aggregation

- ***Example of condition imposed by coherence:***

$(X_1, \hat{p}_1)$ $(X_2, \hat{p}_2)$ $(X_1 \wedge X_2, \hat{p}_3)$ $\hat{p}_1 + \hat{p}_2 - 1 \leq \hat{p}_3 \leq \min\{\hat{p}_1, \hat{p}_2\}$

- ***Forecasts**: Each judge provides forecasts* $\{(\phi_i, \hat{p}_i)\}_{i=1}^{m}$ *for a set of events defined over a common set of basic events* $(X_1, X_2, \ldots, X_n)$

- ***Aggregation:** How do we convert the panel's forecasts in a single, coherent, set of forecasts for the events in question?*

Aggregation/fusion has a rich history

- Within management science, philosophy, psychology, risk analysis, economics, computer science, electrical engineering...
- Behavioral aggregation
  - E.g., Delphi (RAND, Dalkey '67)
- Mathematical aggregation
  - axiomatic, statistical, algorithmic, empirical analyses
  - Bayesian (e.g., Lindley et. al.'79)
  - (a vast literature)
  - Linear averaging (Clemen & Winkler '99)
  - Coherent Approximation Principle (Osherson & Vardi '04)

Linear averaging

	Alice	Bob	Chris	Aggregate
Google ↑	.75	.61	.95	.77
NASDAQ ↓	.2	.1	0	.1

A judge may have insight on non-basic events

- P(Google ↑) = ?

  P(NASDAQ ↓ **AND** Google ↑) ≈ 0.0

- P(Hamas wins) = ?

  P(Hamas loses **OR** Western Aid ↓) ≈ 1.0

- P(Troop Reduction) = ?

  P(Troop Reduction **AND** Insurgency ↑) ≈ 0.0

Linear averaging with non-basic events

	Alice	Bob	Chris	Aggregate
Google ↑	.75	.61	.95	.77
NASDAQ ↓	.2	.1	0	.1
Google ↑ AND NASDAQ ↓	.1	.1	0	.07

- Linear averaging yields coherent aggregate if individual judges are coherent

Averaging with non-basic events (cont'd)

- May be bad to assume/require agents to provide coherent forecasts (e.g., humans, even experts, are **notoriously incoherent** -- Kahneman & Tversky '00)

	Alice	Bob	Chris	Aggregate
Google ↑	.75	.61	.95	.77
NASDAQ ↓	.2	.1	0	.1
Google ↑ AND NASDAQ ↓	**.5**	.1	**.3**	**.3**

$$P(E_1 \cap E_2) \leq \min\{P(E_1), P(E_2)\}$$

- Averaging may lead to an incoherent aggregate.

Incomplete Forecasts

- In practice, forecasts may be "incomplete"

	Alice	Bob	Chris	Aggregate
Google ↑	.75	.61	**NA**	**.68**
NASDAQ ↓	.2	**NA**	0	**.1**
Google ↑ AND NASDAQ ↓	**NA**	.1	.3	**.2**

$$P(E_1 \cap E_2) \le \min\{P(E_1), P(E_2)\}$$

- Again, averaging may lead to an incoherent aggregate.

How to coherently aggregate probability forecasts that may be ***incoherent***, ***incomplete***, *and for* ***logically complex events****?*

Least Squares Projection

(Coherent Approximation Principle – CAP, Osherson & Vardi '05)

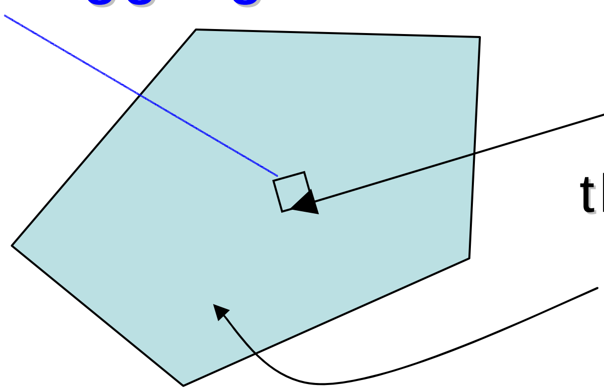
Step 1: Pool all forecasts into single set $\{(\phi_i, \hat{p}_i)\}_{i=1}^{m}$

Step 2: Aggregate by finding the nearest coherent approximation

$$\arg\min \sum_{i=1}^{m} (p_i - \hat{p}_i)^2$$
$$\text{s.t. } \{(\phi_i, p_i)\}_{i=1}^{m} \text{ is coherent}$$

CAP-Aggregate

$\hat{\mathbf{p}}$



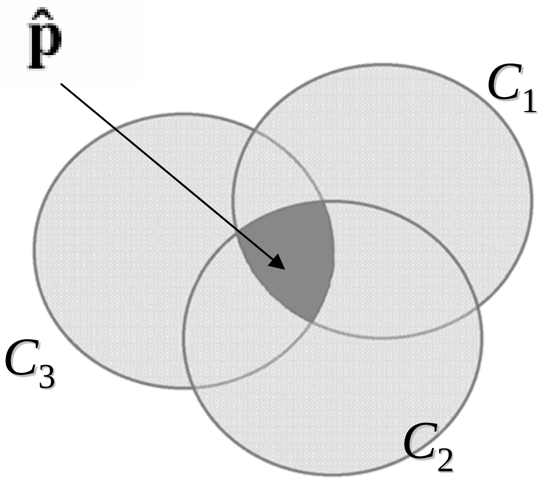
the set of coherent probability assessments for $\{\phi_i\}_{i=1,\ldots,m}$

$\subseteq \mathcal{R}^m$

Algorithmic Issues

- CAP = average when forecasts are coherent but in general is computationally hard.
- linear averaging ~ Local   CAP ~ Global
- Is there something in between, i.e., a scalable algorithm that performs well?

Alternating Projections



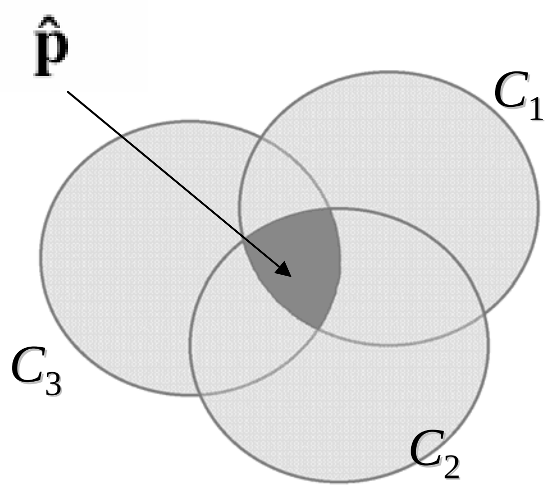
$$P_C(\hat{\mathbf{p}}) = \arg\min \|\mathbf{p} - \hat{\mathbf{p}}\|_2^2$$
$$\text{s.t.} \quad \mathbf{p} \in C = \cap_{i=1}^{m} C_i$$

von Neumann-Halperin [Vo50]

$$\mathbf{p}_0 = \hat{\mathbf{p}}$$
$$\mathbf{p}_n = P_{C_{(n \bmod m)+1}}(\mathbf{p}_{n-1}) \qquad n = 1, 2, \ldots$$

Alternating Projections (cont'd)

Theorem: Under regularity,



- von Neumann-Halperin converges
- $\lim_{n\to\infty} \mathbf{p}_n \in \cap_{j=1}^{m} C_i$
- if C_i affine, $\lim_{n\to\infty} \mathbf{p}_n = P_C(\hat{\mathbf{p}})$
- $\|\mathbf{p}_n - \mathbf{p}\|_2 \le \|\mathbf{p}_{n-1} - \mathbf{p}\|_2, \quad \forall \mathbf{p} \in \cap_{j=1}^{m} C_j$

A General Approach – Enforce Local Coherence Constraints

Input: a set of forecasts $\{(\phi_i, \hat{p}_i)\}_{i=1}^m$

Step 1: decompose the events $\{\phi_i\}_{i=1}^m$ into subsets of events $\{C_j\}_{j=1}^l$ s.t $C_j \subseteq \{\phi_i\}_{i=1}^m$

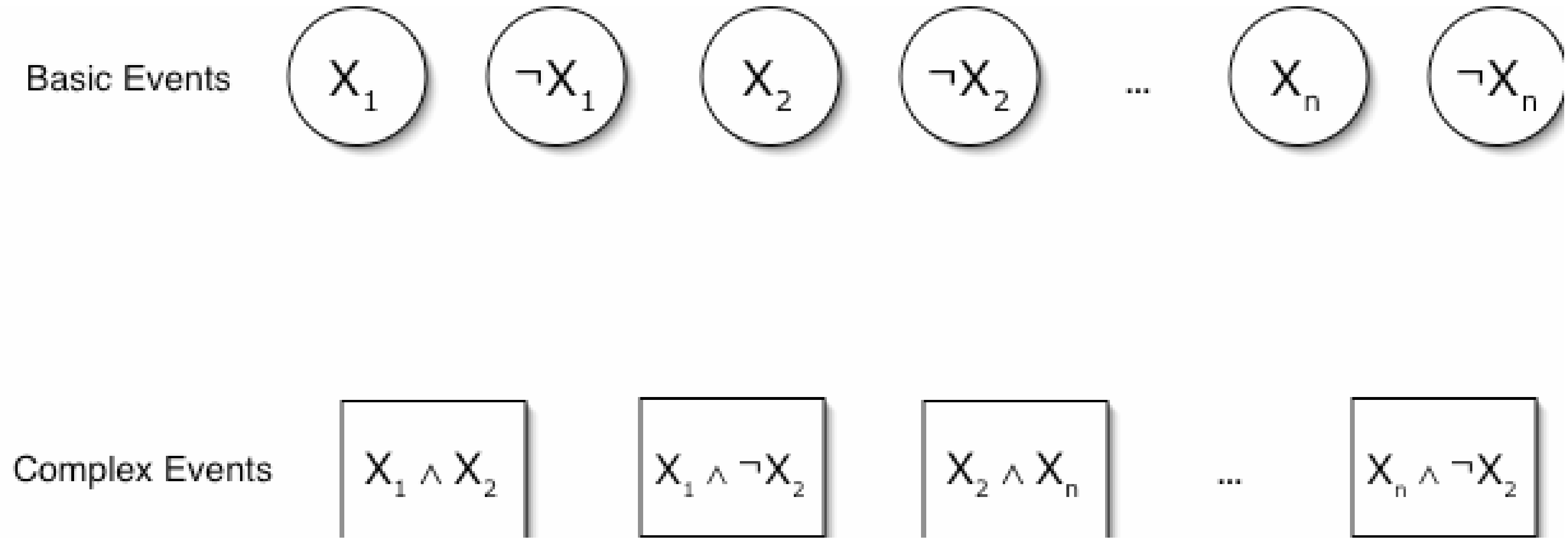
$\{(\phi_i, \hat{p}_i)\}_{i=1}^m$ is **locally coherent with respect to C$_j$** iff $\{(\phi_i, p_i)\}_{\phi_i \in C_j}$ is probabilistically coherent

Step 2: Solve a relaxation of the CAP:

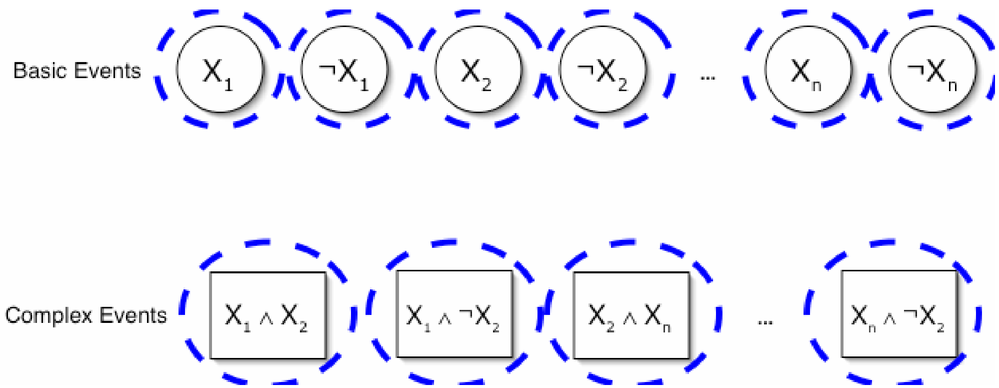
$$\min \sum_{i=1}^m (p_i - \hat{p}_i)^2$$
$$\text{s.t. } \{(\phi_i, p_i)\} \text{ is locally coherent wrt to } C_j \ \forall j = 1, \ldots, l$$

Use **von Neumann-Halperin**

Local Coherence Constraints?

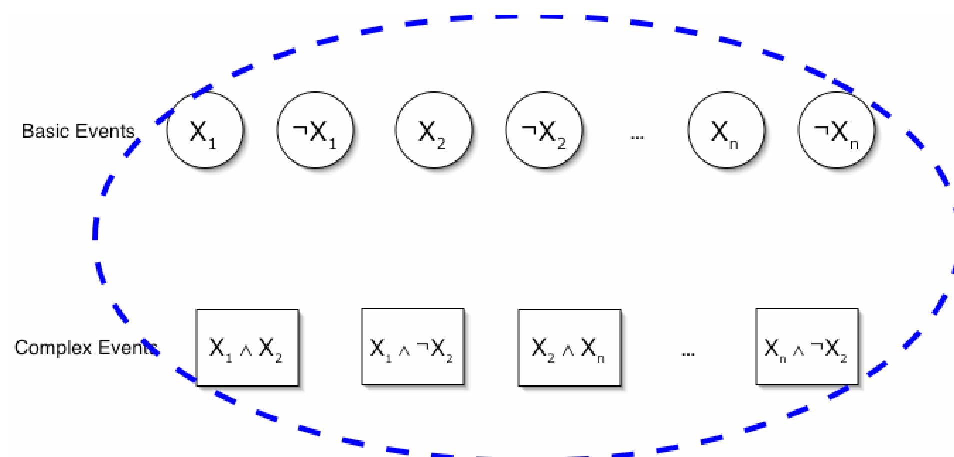


Local Coherence Constraints?
Degenerate special cases

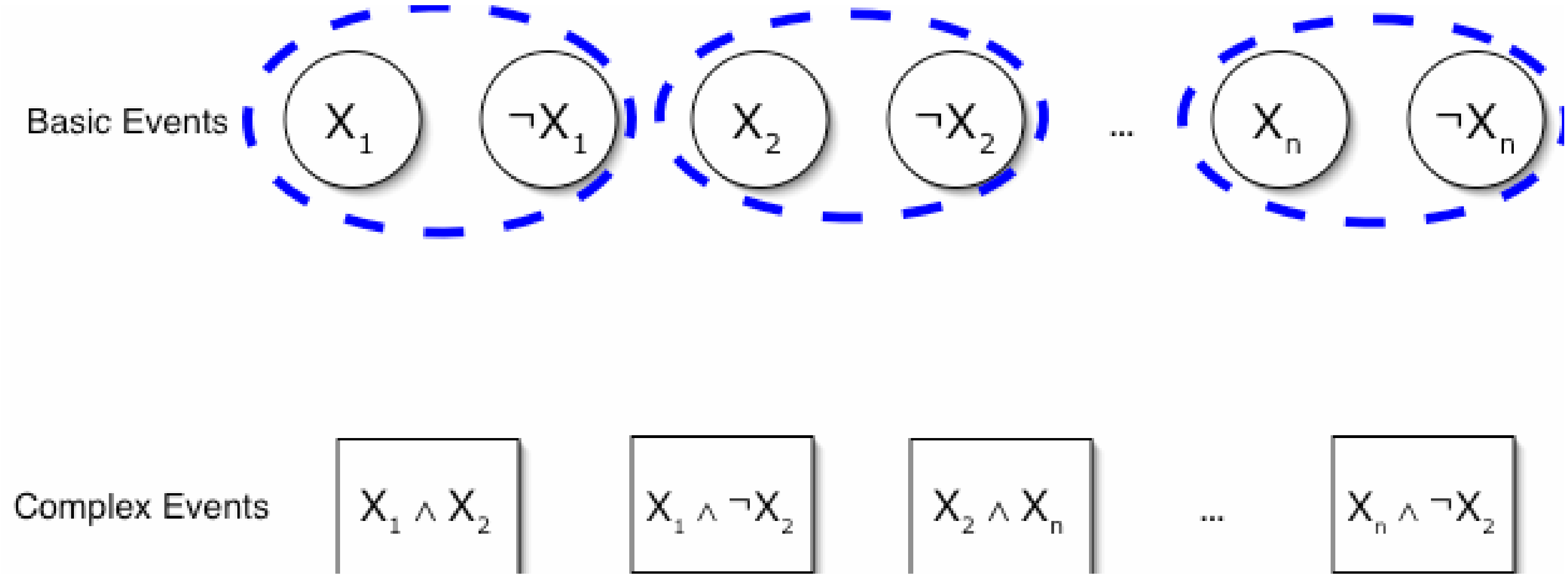


Linear Averaging

CAP

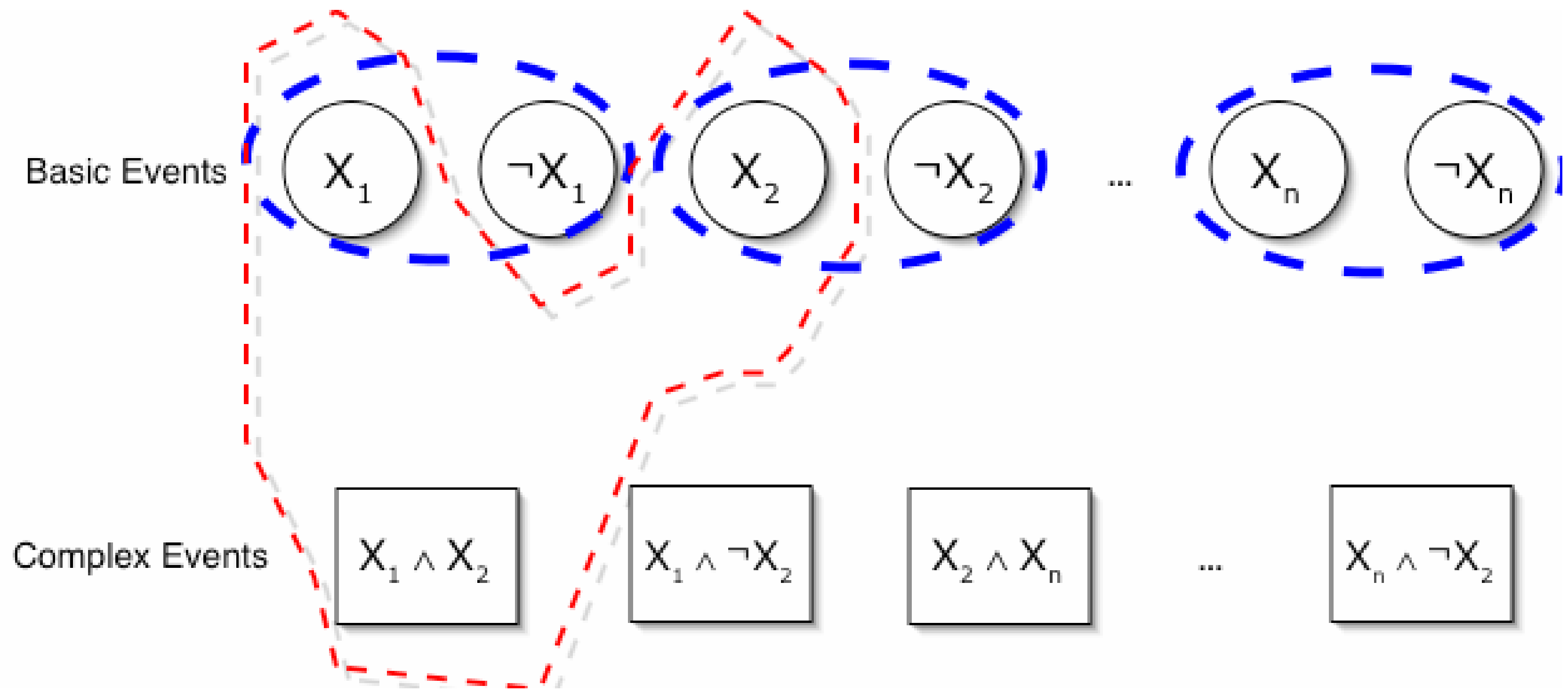


Local Coherence Constraints?
An Example



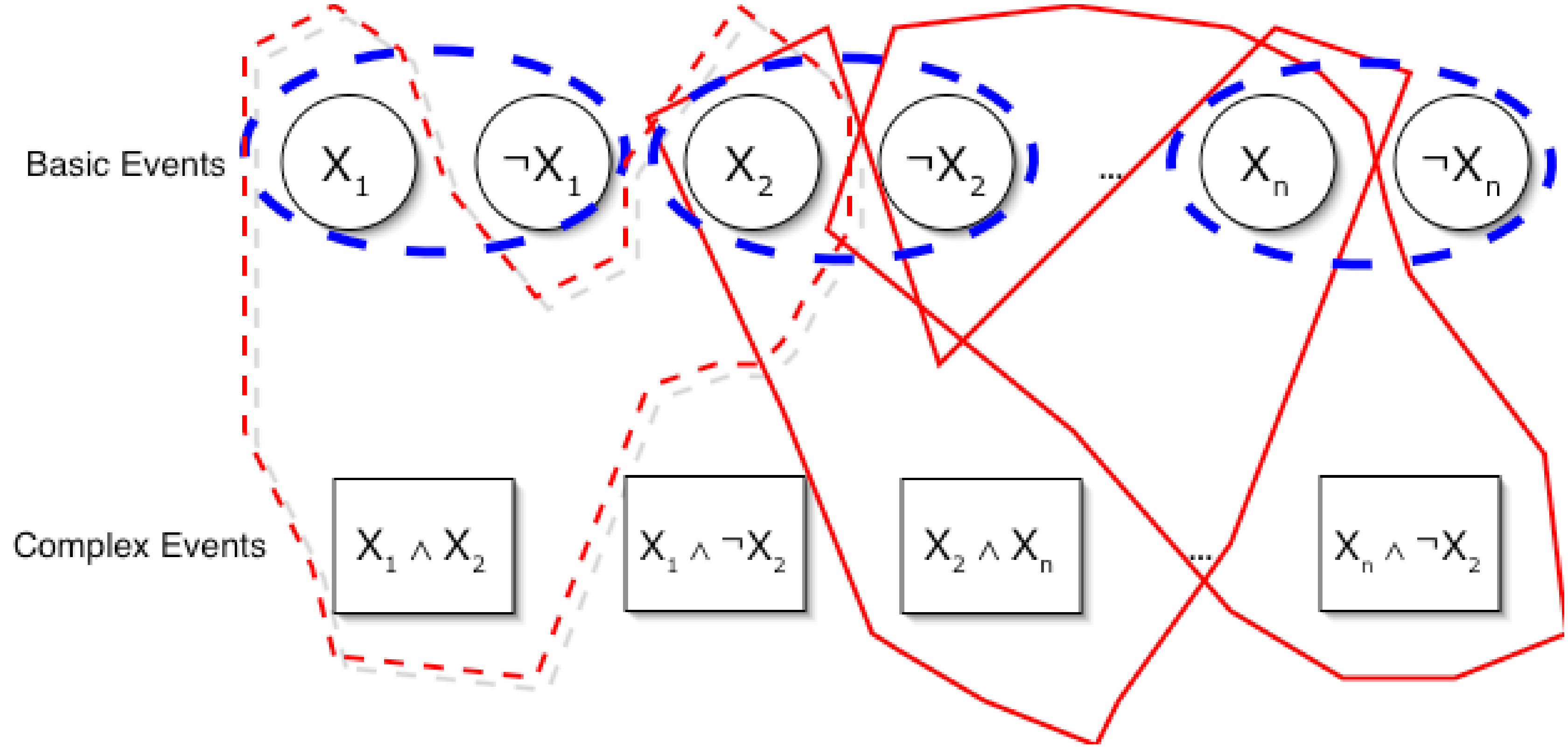
Local Coherence Constraints?

An Example



Local Coherence Constraints?

An Example



How to choose the subsets?

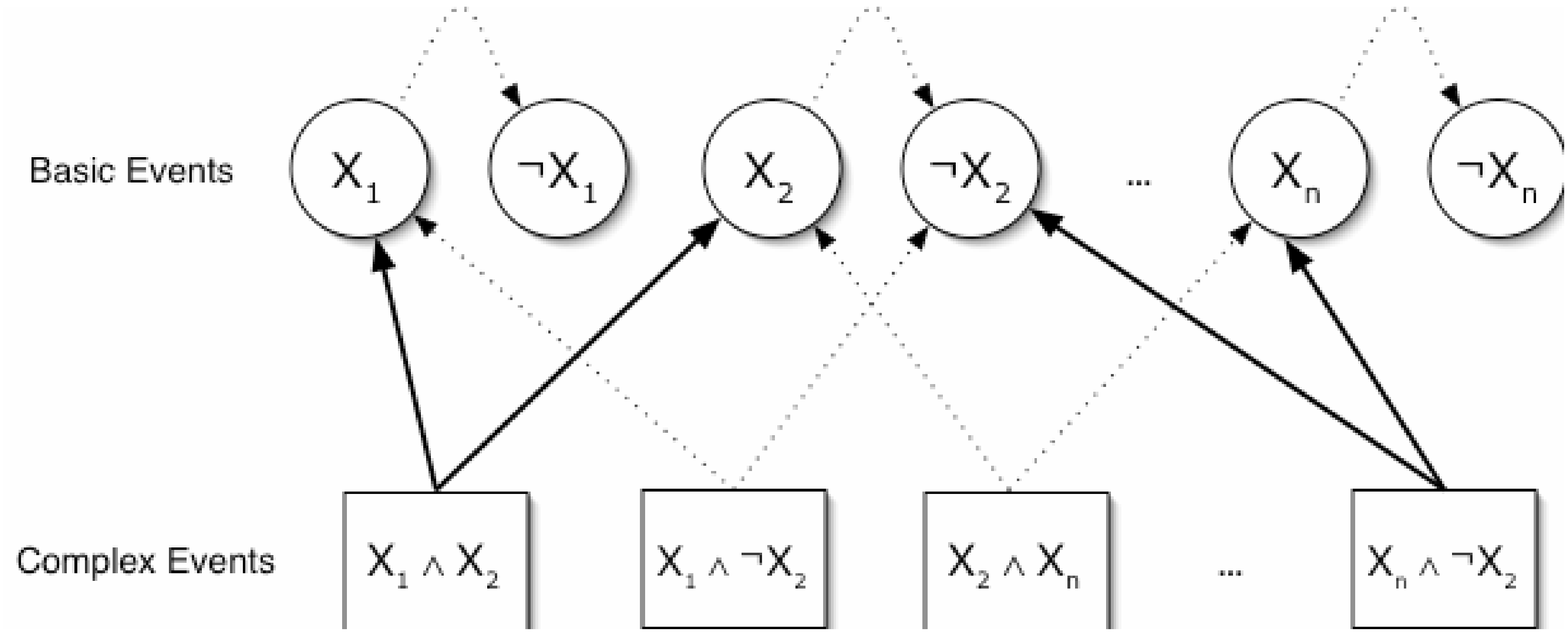
- *Larger subsets, closer approximation to CAP-aggregate*

- *Smaller subsets, fast aggregation*

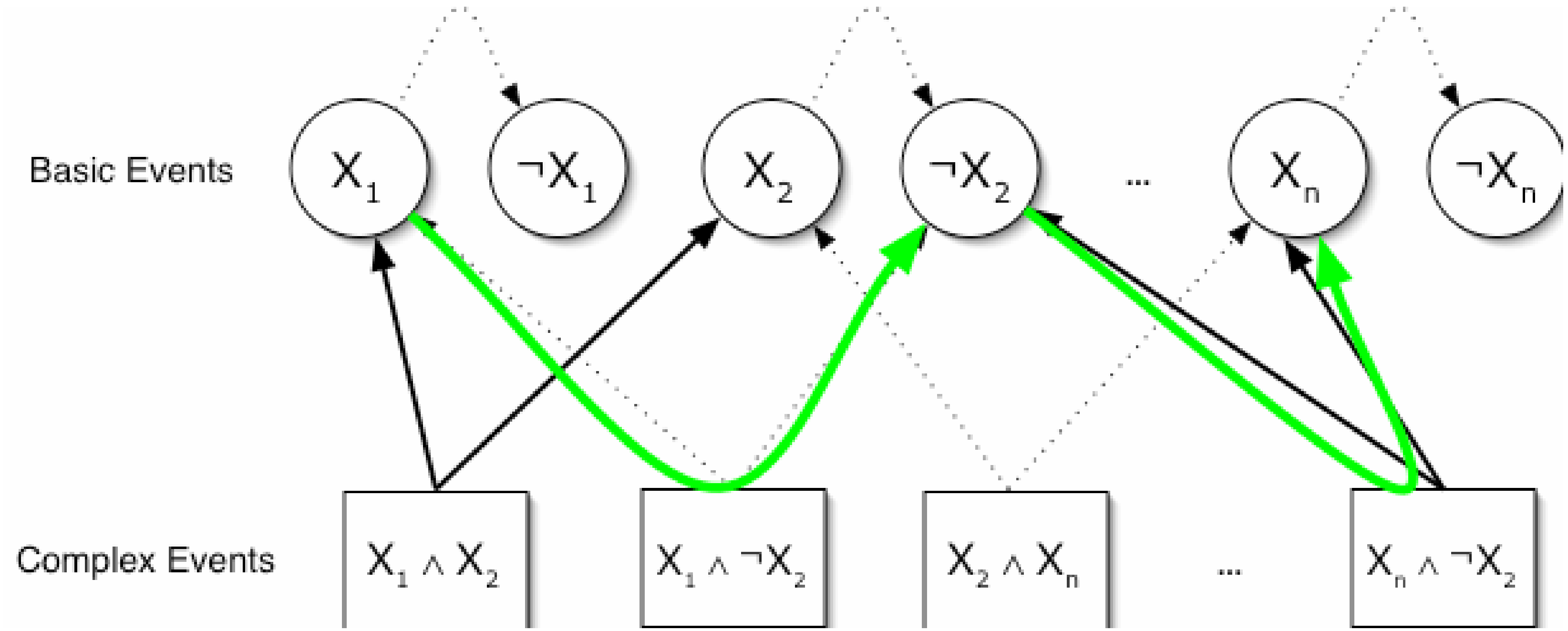
- *Exploit logical simplicity to trade-off speed vs. approximation*

- *Overlapping subsets - "global information propagation"*

A Graph Structure



Local information propagates globally



A Performance Criterion

- Suppose that after learning the **truth**, we measure the accuracy of a set $\{(\phi_i, \hat{p}_i)\}_{i=1}^{m}$ of forecasts using the **Brier Score** (Brier '50)

$$\sum_{i:\phi_i=\text{TRUE}} (1-\hat{p}_i)^2 + \sum_{i:\phi_i=\text{FALSE}} (0-\hat{p}_i)^2$$

A Performance Guarantee

Theorem:

Let $\{(\phi_i, \hat{p}_{T,i})\}_{i=1}^{m}$ denote the aggregate forecasts after T iterations. Then,

- The aggregate **Brier score improves at each step**: $\text{Brier}(T) \leq \text{Brier}(T-1)$
- and as $T \to \infty$, $\{(\phi_i, \hat{p}_{T,i})\}_{i=1}^{m}$ is locally coherent w.r.t. all designed subsets.

Proof: von Neumann-Halperin +
Generalized de Finetti Theorem

Five Data Sets

- *FINANCE (Batsell et. al., 2002)*
  - *Performance of 10 Stocks in 3rd Quarter 2000*
- *STOCKS (Osherson & Vardi, 2005)*
  - *Economic Indicators in 4th Quarter 2001*
- *NBA1 , NBA2 (Batsell et. al., 2002)*
  - *Two Nat'l Basketball Association (NBA) games*
- *HOUSTON (Hendrix et. al., 2005)*
  - *Pollution Levels & Real Estate Prices in Houston*

Sample questions

What is the probability that

United Airlines outperforms the S&P500?

Please estimate the probability that on Monday, July 22nd:

the temperature will exceed 85 degrees at Hobby airport

AND

there will NOT be an ozone watch in the ship channel

What is the probability that

IBM stock price increases

OR

Dell stock price increases

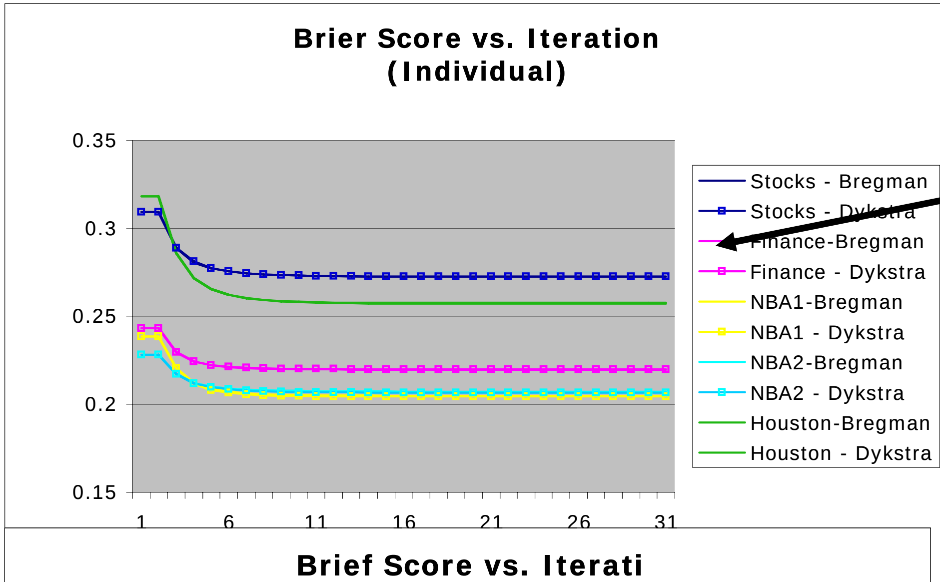
Five Data Sets

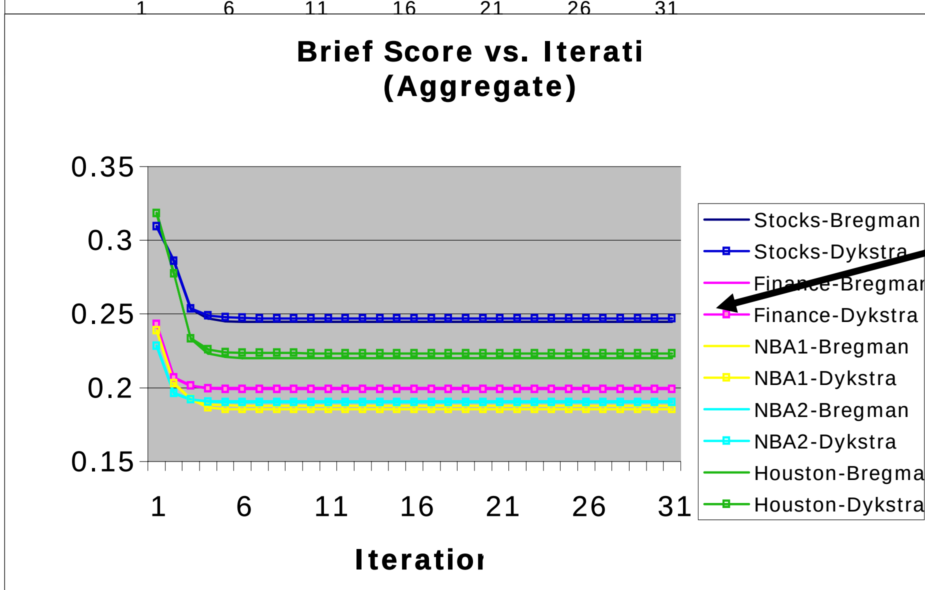
- **34 Logically Simple (10) & Complex Events (24) per subject**

$p \quad p \wedge q \quad p \vee q \quad p \wedge \neg q \quad p \vee \neg q$

	STCK	FIN	NBA1	NBA2	HSTN
Subjects	47	31	29	36	17
Basic Events	30	10	10	10	10
Events/Aggregate	1598	1054	986	1224	578

How **fast**? How does it **scale**?





(34 Events/subject)

~**.6s**

1598 Events/aggregate
(1000 unique events)

~**10s**

SAPA: Multiple hours

How does aggregation impact forecasting accuracy?

		Stocks	Finance	NBA1	NBA2	Houston
Brier Score	RAW	0.309	0.243	0.239	0.228	0.318
	Individual	0.273	0.220	0.205	0.207	0.257
	Aggregate	0.245	0.200	0.188	0.191	0.220
	Linear Average	0.286	0.207	0.203	0.196	0.234
Slope	RAW	0.064	0.153	0.140	0.141	0.129
	Individual	0.109	0.172	0.186	0.169	0.210
	Aggregate	0.114	0.153	0.173	0.150	0.202

SAPA: 0.276

"Optimal" CAP: 0.272

Summary

- Fast algorithm for aggregating incoherent, incomplete forecasts for complex events.
- Connections with collaborative regression and other areas.
- Next steps?
  - Probability aggregation as an approach for hard/soft fusion, distributed inference, etc.?
  - Better algorithms, analysis, applications
  - Performance criteria?
  - Dealing with other types of forecasts (conditional, binary, relative, confidence bounds, etc.)

Thank you!