Clustering Prediction of Opinion Dynamics

Fei Lu
Department of Mathematics
Johns Hopkins University

Joint with **Zehong Zhang**
@ Dept. Math, JHU

2021.07.19 SIAM Annual Meeting

Uncertainty Quantification Strategies for Data-Driven, Large-Scale Problems

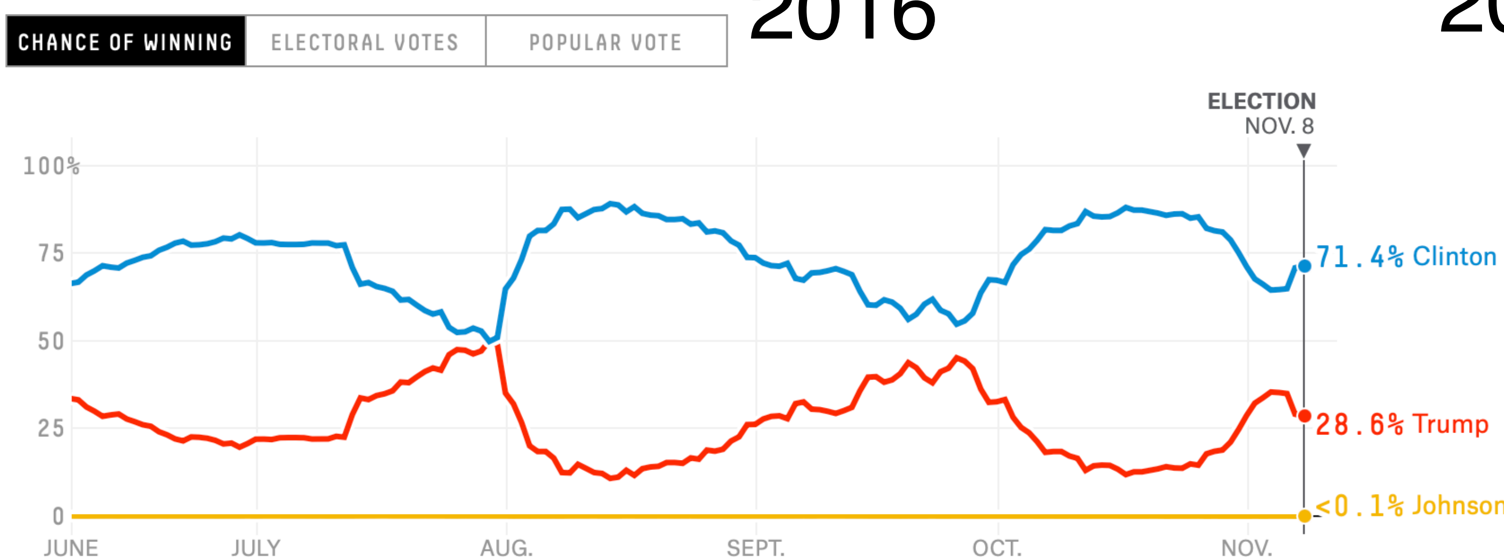
Research supported by JOHNS HOPKINS UNIVERSITY NSF

Motivation and problem statement

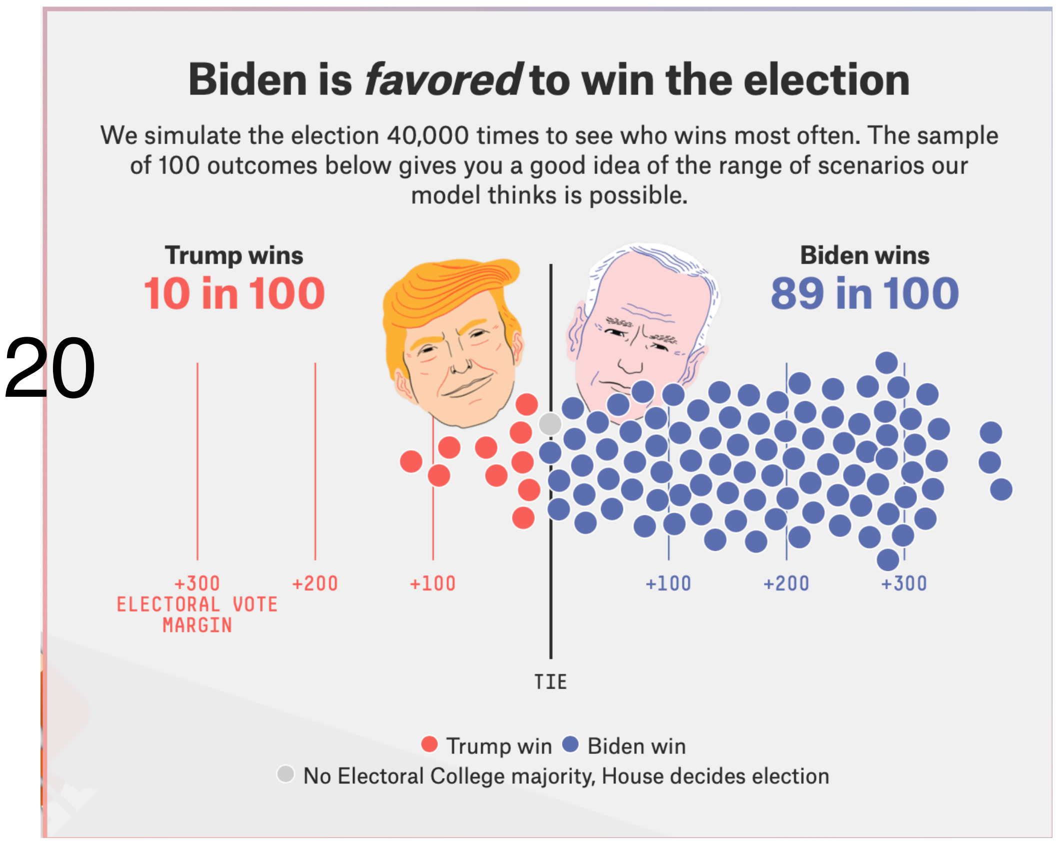
Political election: difficult to predict

538 predictions of US president election

2016



2020



ELECTIONS

Pollsters: 'Impossible' to say why 2020 polls were wrong

A new report couldn't answer the big question plaguing political polling: Why were surveys off by so much in 2020?

Fundamental elements in political election: **Opinions**

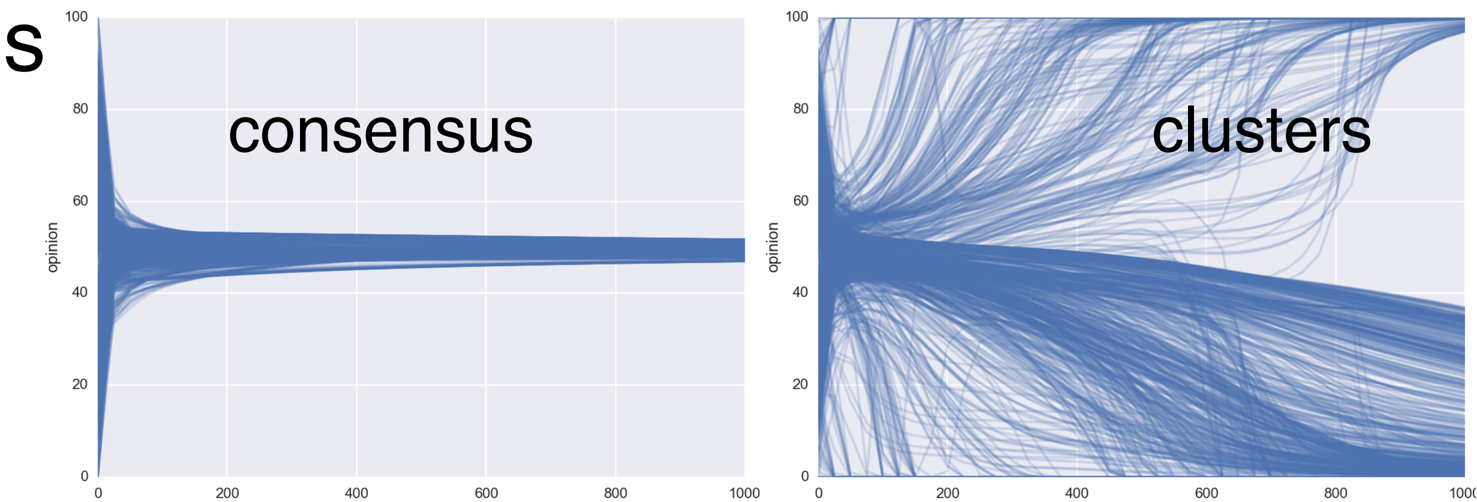
Problem statement

Opinions evolve. How to predict them?

- **Describe the dynamics:**

"opinion dynamics", "agent-based models", "interacting particles"

- discrete- /continuous- models [Krause 2000, Motsch+Tadmor 2014,Duggins17…]
- **clusters** / consensus



Duggins17JASSS

- **Dynamics + Data**

- learn the dynamics from data
- **partial noisy data:** uncertainty >>> **Bayesian prediction**

>> Prediction, Control / influence

Outline

Problem statement

1. State-space model formulation

2. Bayesian approach for UQ
>>> sample the posterior ↓

3. Sequential Monte Carlo
>> auxiliary implicit sampling

4. Numerical example

5. Randomization enhances observability

Summary

1. State-space model formulation

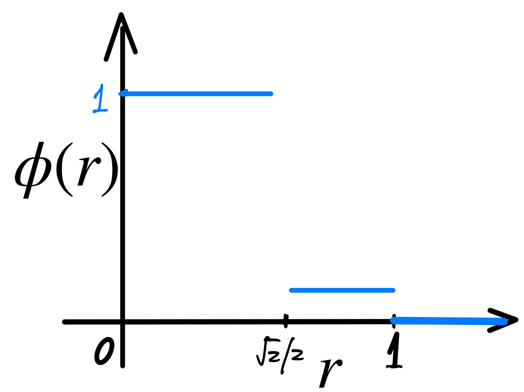
Opinion Dynamics

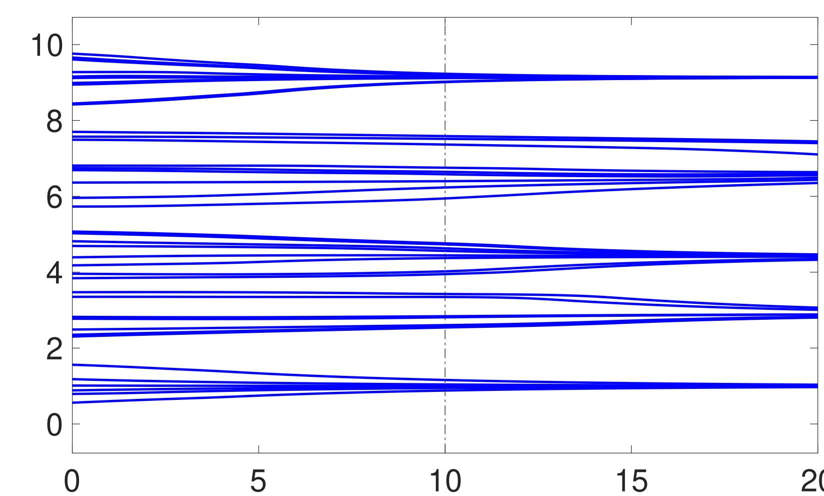
$$x_{t+1}^i - x_t^i = \frac{1}{N}\sum_{j=1}^{N} \phi(|x_t^j - x_t^i|)(x_t^j - x_t^i)\Delta t$$

x_t^i — opinion of agent i at time t (in $\mathbb{R}^d$, d arbitrary)

N — number of agents

Local interaction





Clusters:
- disjoint sets in a partition of the agents
- invariant once emerge:

cluster size $|\mathcal{C}_k| := |\mathcal{C}_k(t)| = |\mathcal{C}_k(t_c)|,$

cluster center $\overline{x}_{\mathcal{C}_k} := \frac{1}{|\mathcal{C}_k(t)|}\sum_{i\in\mathcal{C}_k(t)} x_t^i = \frac{1}{|\mathcal{C}_k(t)|}\sum_{i\in\mathcal{C}_k(t)} x_{t_c}^i.$

1. State-space model formulation

Dynamics + Data >>> state-space model

$$\begin{cases} x_{t+1} = g(x_t), \;\; x_1 \sim \mu(\cdot), & \text{state model} \\ z_t = Hx_t + \xi_t, & \text{observation model} \end{cases}$$

- data $z_{1:T}$; initial distribution μ known;
- observe $N_1 < N$ agents: H is a projection;
- nonlinear state model; Gaussian noise

Observability (noiseless data, linear system)

> Theorem [ZhangLu20] a linear system is observable iff at most 1 agent is not observed

Bayesian: **posterior of states and clusters**

2. Bayesian approach

Bayesian: **posterior of states and clusters**

$p(x_{1:T} \mid z_{1:T})$,	posterior of $x_{1:T}$ conditional on $z_{1:T}$
$\widehat{p}(x_{1:T} \mid z_{1:T})$	empirical approximation of $p(x_{1:T} \mid z_{1:T})$
$\{x_{1:t}^{(s)}, w_t^{(s)}\}$	samples and weights
$p(\lvert\mathcal{C}_i\rvert \mid z_{1:T})$, $p(\overline{x}_{\mathcal{C}_i} \mid z_{1:T})$	posteriors of $\lvert\mathcal{C}_i\rvert$ and $\overline{x}_{\mathcal{C}_i}$

Samples $\{x_T^{(s)}, w_T^{(s)}\}$ $\xrightarrow{\text{state model}}$ samples of clusters

Sampling the posterior

- high dimension: MCMC no good
- **Sequential Monte Carlo**: sequential importance sampling

  [Doucet+Johansen2009, Liu2008,…]
- Particle MCMC: SMC+MCMC [Andrieu etc2010, Lindsten etc2014]

3. Sequential Monte Carlo

Sequential importance sampling:

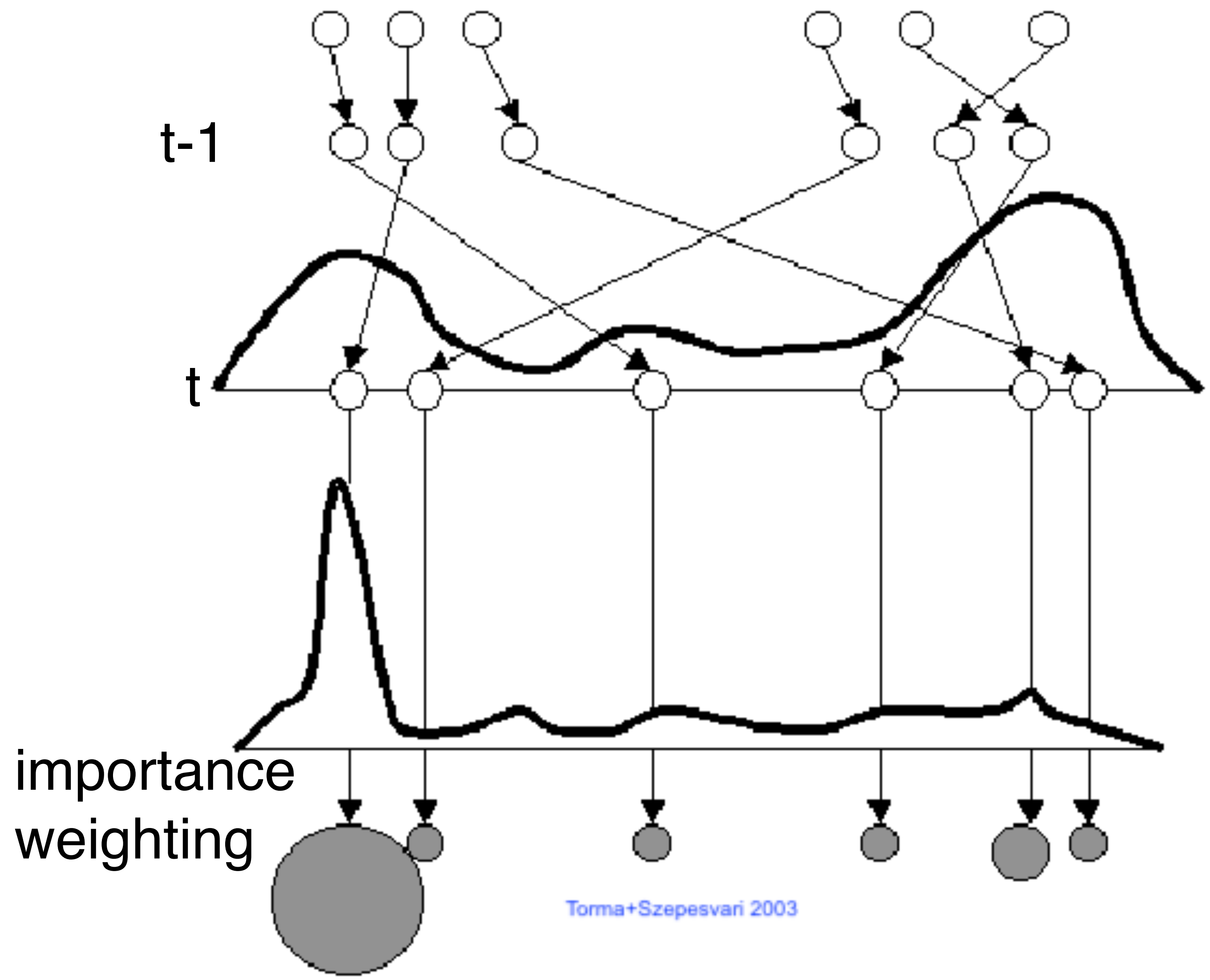
$$p(x_{1:t} | z_{1:t}) = p(x_{1:t-1} \mid z_{1:t-1}) \frac{p(x_t | x_{t-1}) p(z_t \mid x_t)}{p(z_t | z_{1:t-1})},$$

Weighted samples from importance density

$$x_t^{(s)} \sim q(x_t | x_{1:t=1}^{(s)}, z_{1:t})$$

$$w_t^{(s)} = w_{t-1}^{(s)} \cdot \frac{p(z_t \mid x_t^{(s)}) \cdot p(x_t^{(s)} \mid x_{t-1}^{(s)})}{q(x_t^{(s)} \mid x_{1:t-1}^{(s)}, z_{1:t})}.$$

- recursive for all t;
- resampling to reduce degeneracy
- **Key: good importance densities**



3. Sequential Monte Carlo

$$\begin{cases} x_{t+1} = g(x_t) + \epsilon_t, \quad x_1^i \sim \mu, \forall i, \\ z_t = Hx_t + \xi_t, \end{cases}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2 I_{dN})$ and $\xi_t \sim \mathcal{N}(0, \sigma_\xi^2 I_{dN_1})$.

Implicit sampling: [Chorin, Tu, Morzfel] optimal 1-step importance sampling

$$p(x_{1:t} \mid z_{1:t}) = p(x_{1:t-1} \mid z_{1:t-1}) \boxed{\frac{p(x_t \mid x_{t-1}) p(z_t \mid x_t)}{p(z_t \mid z_{1:t-1})}}, \longrightarrow q^{\text{opt}}(x_t \mid x_{1:t-1}, z_{1:t})$$

- Gaussian bc linear Gaussian observation model
- **not updated using information from new observations**

Auxiliary implicit sampling:

- using two-step observations (look ahead strategy) [Pit+Shephard1999,Lin+Chen+Liu2013,…]
- linear approximation of the state model
  >>> informative importance density

4. Numerical results: State estimation

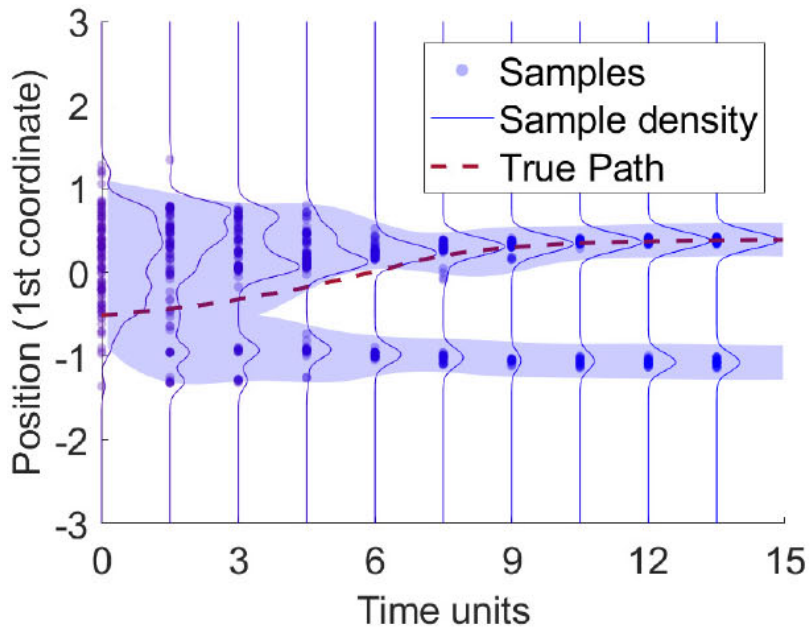
N1= 30, N=60, d=2

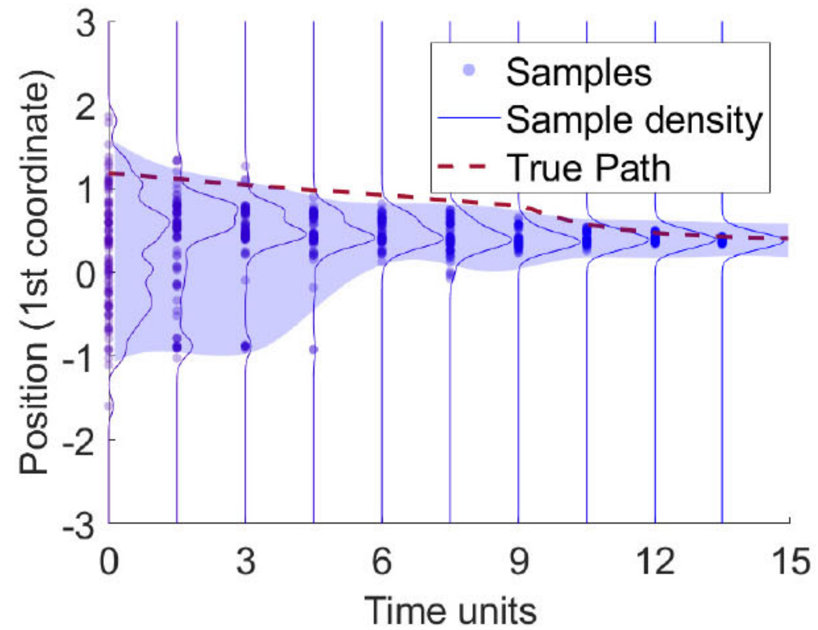
- multi-mode posterior

Marginal posterior of x_t^1

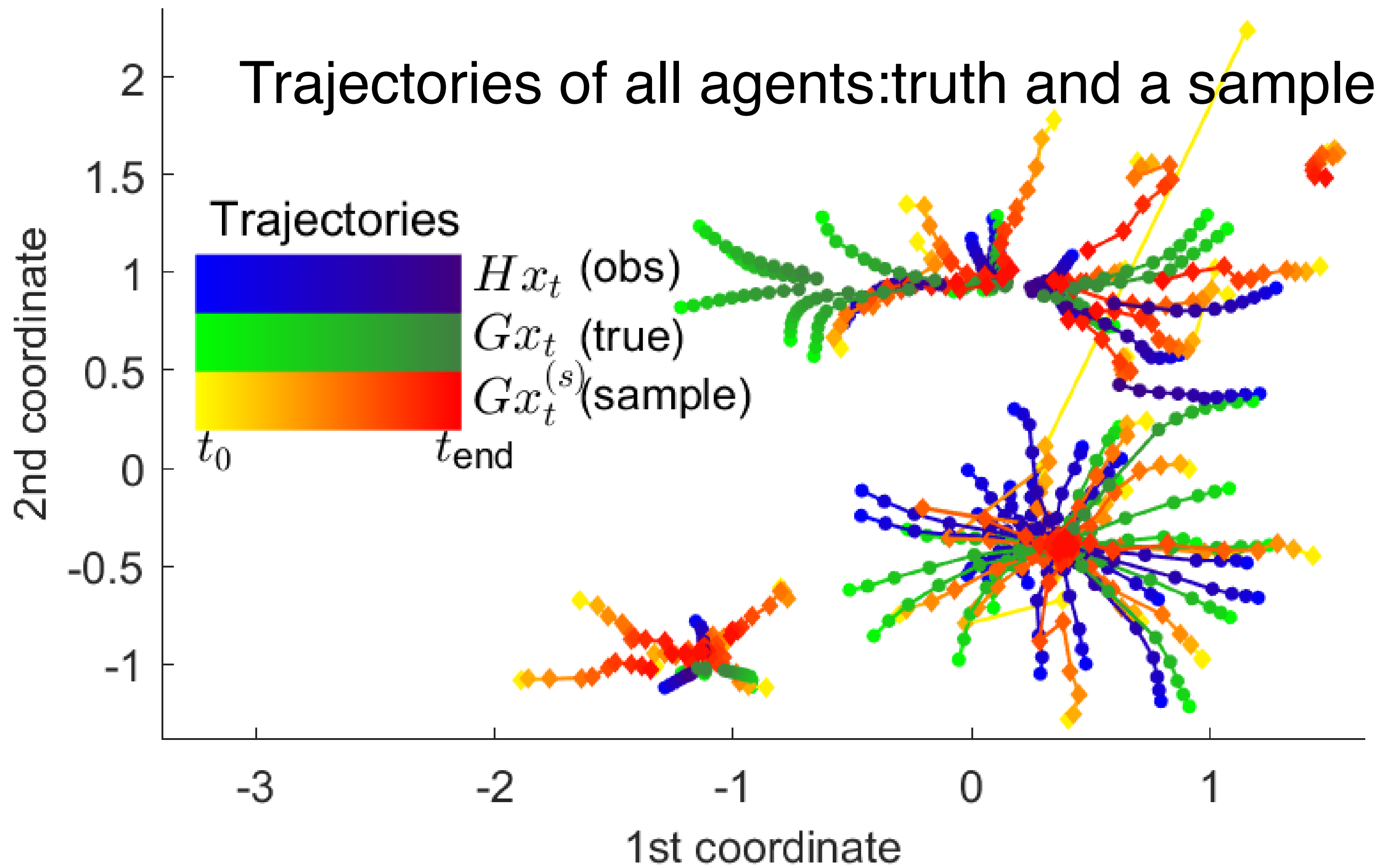
(of two agents,100 samples)

  - shaded: 95% CI





- sample path ~= truth
- the clustering is close



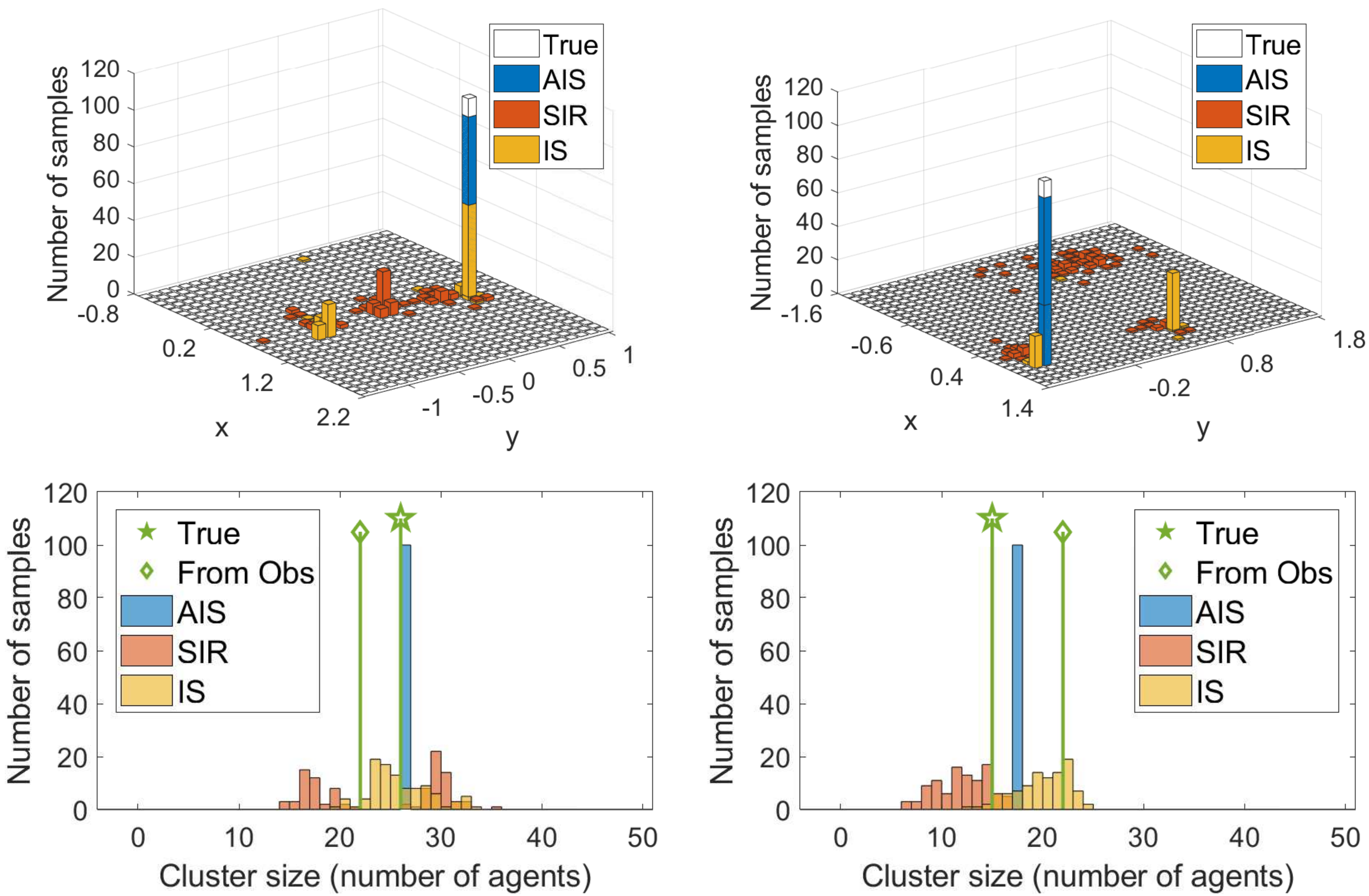
4. Numerical results: cluster prediction

Posterior of cluster's centers and sizes AIS > IS > SIR

the largest cluster | the 2nd largest cluster

Center:

Size:



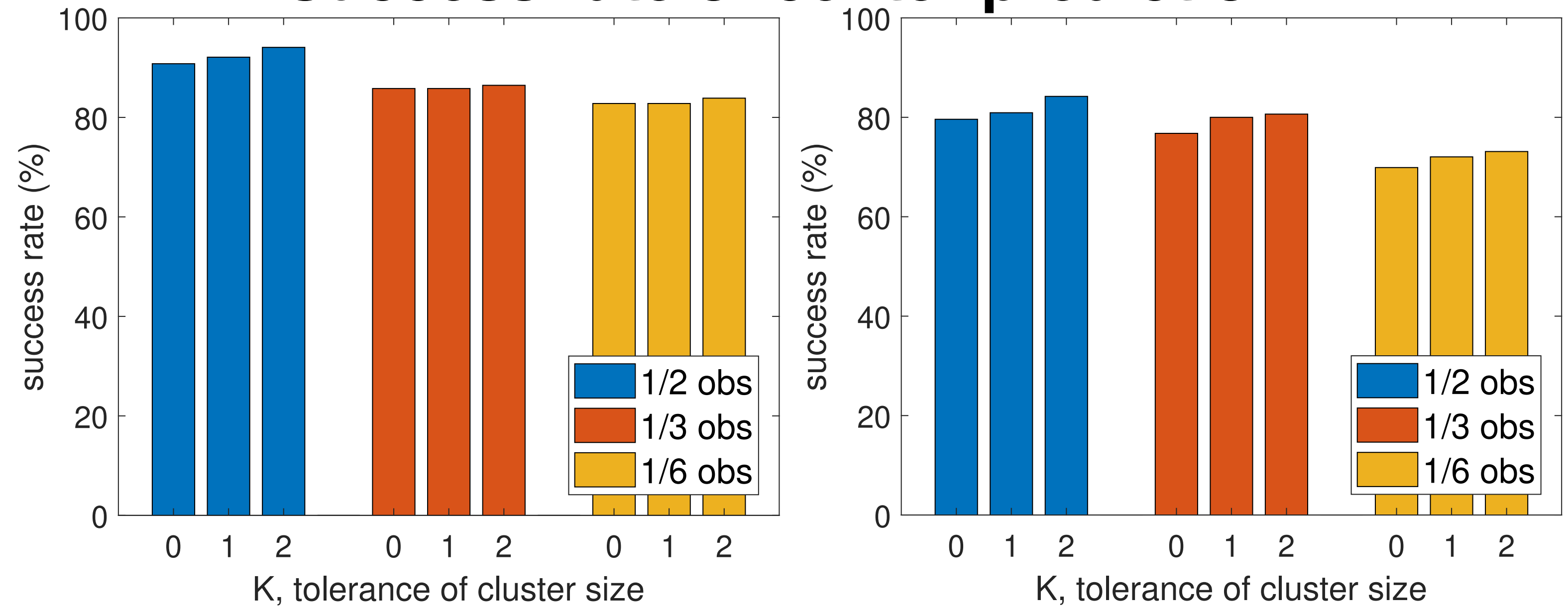
4. Numerical results: cluster prediction

How robust is the predictions in 100 simulations ?

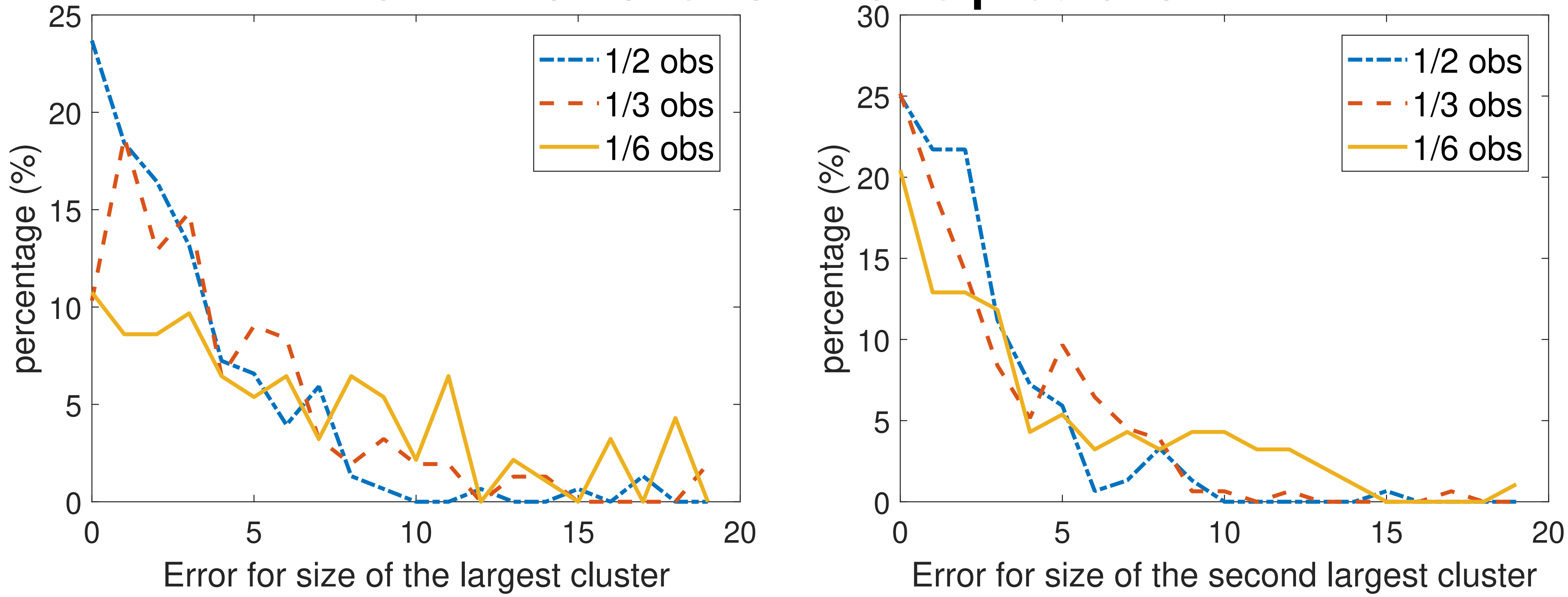
Center: ~90% success

Size: difficult to predict

Success rate of center prediction



Distribution of error in size prediction



5. Randomization enhances observability

Instead of tracking, randomly poll at each time

Observability (noiseless data, linear system)

Theorem (Lu-Zhang-Zhang21): randomly pick K out of N at each time. At time T, the linear system is observable with probability

$$\mathbb{P}\,(\text{all but one is visited}) \approx 1 - N(1-\frac{K}{N})^T$$

- exponential goes to 1 as T increases
- observation ratio K/N in the base

$$\mathbb{P}\,(\text{visit all but one}) = N(1-\tfrac{K}{N})^T \left[1 - \sum_{\ell=1}^{N-1-K} (-1)^{\ell-1} \binom{N-1}{\ell} \prod_{j=1}^{\ell} (1-\tfrac{K}{N-j})^T \right]$$

Summary

1. Bayesian prediction for clusters of opinion dynamics

2. Auxiliary implicit sampling (AIS) SMC method

- two-step observations (a lookahead strategy)

3. Randomization enhances observability

Open research

- Prediction for stochastic systems
- Privacy-preserving randomization

References

[1] Zehong Zhang and FL (2020) Transactions on Signal and Information Processing over Networks.

[2] FL, Cheng Zhang and Zehong Zhang (2021) (on-going) Randomization enhances observability of multi-agent systems