

Learning From Glauber Dynamics

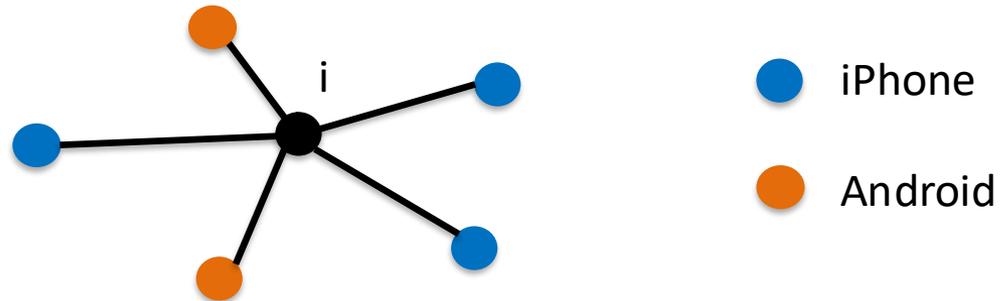
Ankur Moitra (MIT)

Swiss Winter School, Lecture #3

How do Ising models arise in social networks?

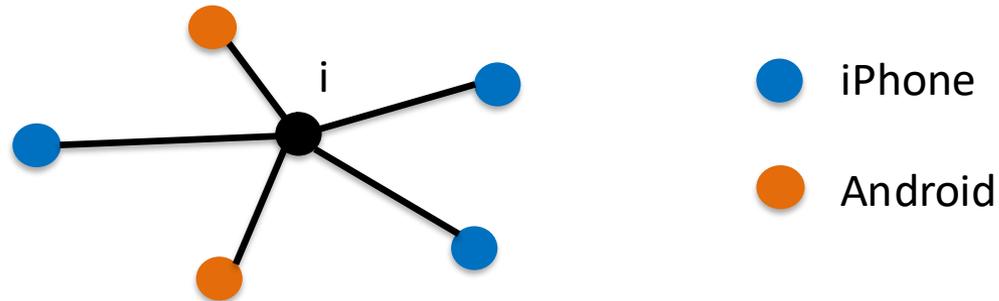
RUNNING EXAMPLE

Modeling the adoption of technology in a social network



RUNNING EXAMPLE

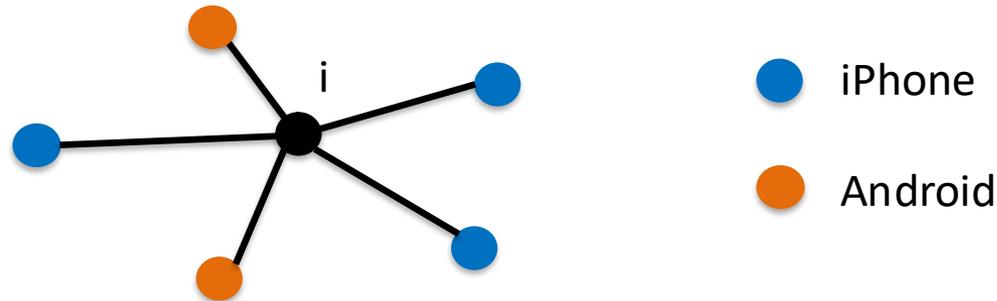
Modeling the adoption of technology in a social network



Many works formalize as an Ising model

RUNNING EXAMPLE

Modeling the adoption of technology in a social network



Many works formalize as an Ising model

Each person's decision is independent of everyone else,
conditioned on his neighbors

Main takeaway from Lecture #1:

Proposition: Can learn bounded degree Ising models at any temperature in polynomial time, but for Markov random fields of order k need n^k time

Main takeaway from Lecture #1:

Proposition: Can learn bounded degree Ising models at any temperature in polynomial time, but for Markov random fields of order k need n^k time

But where would these iid samples come from?

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

- An Algorithmic Surprise
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- Taming Higher-Order Interactions

OUTLINE

Part I: Introduction and Review

- **Computational Lower Bounds**

Part II: Learning from Trajectories

- An Algorithmic Surprise
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- Taming Higher-Order Interactions

COMPLEXITY OF SAMPLING

Theorem (informal): The Glauber dynamics has polynomial mixing time if $\beta < \beta_c$

COMPLEXITY OF SAMPLING

Theorem (informal): The Glauber dynamics has polynomial mixing time if $\beta < \beta_c$

Above the critical temperature Glauber mixes slowly, even worse:

Theorem [Sly '10], [Sly, Sun '12]: It's not just that Glauber is slowly mixing when $\beta > \beta_c$, but in fact sampling is NP-hard

Main takeaway from Lecture #1:

Proposition: Can learn bounded degree Ising models at any temperature in polynomial time, but for Markov random fields of order k need n^k time

But where would these iid samples come from?

Main takeaway from Lecture #1:

Proposition: Can learn bounded degree Ising models at any temperature in polynomial time, but for Markov random fields of order k need n^k time

But where would these iid samples come from?

Often NP-hard

Main takeaway from Lecture #1:

Proposition: Can learn bounded degree Ising models at any temperature in polynomial time, but for Markov random fields of order k need n^k time

But where would these iid samples come from?

Often NP-hard

A quip due to Kamal Jain:

“If my laptop can’t find it, then neither can the market”

Main takeaway from Lecture #1:

Proposition: Can learn bounded degree Ising models at any temperature in polynomial time, but for Markov random fields of order k need n^k time

But where would these iid samples come from?

Often NP-hard

A quip due to Kamal Jain:

“If my laptop can’t find it, then ~~neither can the market~~”

it’s not a defensible generative model

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

- An Algorithmic Surprise
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- Taming Higher-Order Interactions

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

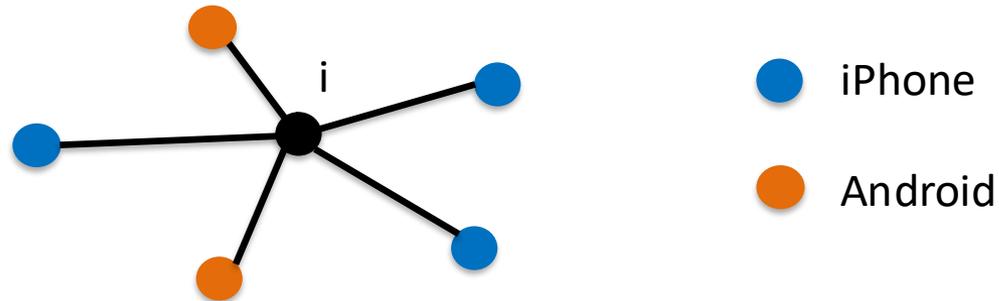
- An Algorithmic Surprise
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- Taming Higher-Order Interactions

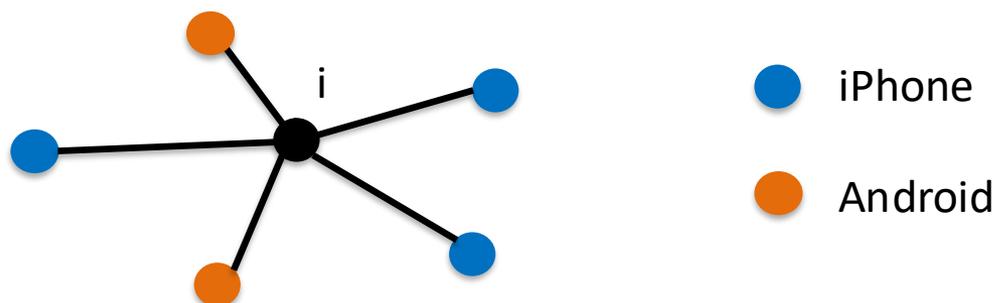
RUNNING EXAMPLE, REVISITED

In fact, for adoption of technology on a social network...



RUNNING EXAMPLE, REVISITED

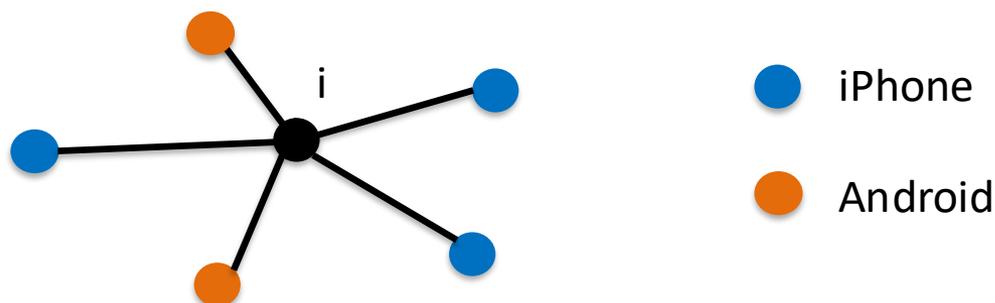
In fact, for adoption of technology on a social network



Many works [**Montanari, Saberi '10**] model the time-dependent process as Glauber dynamics

RUNNING EXAMPLE, REVISITED

In fact, for adoption of technology on a social network



Many works **[Montanari, Saberi '10]** model the time-dependent process as Glauber dynamics

LEARNING WITHOUT IID SAMPLES

Even when Glauber doesn't mix, can we still learn from it?

LEARNING WITHOUT IID SAMPLES

Even when Glauber doesn't mix, can we still learn from it?

[Bresler, Gamarnik, Shah '15]: Polynomial time algorithm for structure learning for Ising models from Glauber dynamics*

LEARNING WITHOUT IID SAMPLES

Even when Glauber doesn't mix, can we still learn from it?

[Bresler, Gamarnik, Shah '15]: Polynomial time algorithm for structure learning for Ising models from Glauber dynamics*

[Gaitonde, Mossel '25]: Can do parameter learning too*

LEARNING WITHOUT IID SAMPLES

Even when Glauber doesn't mix, can we still learn from it?

[Bresler, Gamarnik, Shah '15]: Polynomial time algorithm for structure learning for Ising models from Glauber dynamics*

[Gaitonde, Mossel '25]: Can do parameter learning too*

*Assumes we see not just when nodes switch, but also when they considered switching and didn't

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

- An Algorithmic Surprise
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- Taming Higher-Order Interactions

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

- **An Algorithmic Surprise**
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- Taming Higher-Order Interactions

AN ALGORITHMIC SURPRISE

Instead of hoping to get the same results for learning from Glauber as is possible from i.i.d. samples....

AN ALGORITHMIC SURPRISE

Instead of hoping to get the same results for learning from Glauber as is possible from i.i.d. samples....

Is learning from Glauber dynamics *easier*?

AN ALGORITHMIC SURPRISE

Instead of hoping to get the same results for learning from Glauber as is possible from i.i.d. samples....

Is learning from Glauber dynamics *easier*?

Theorem [Gaitonde, Moitra, Mossel '24]: There is an algorithm that learns Markov random fields of order k in $O(n^2 \log n)$ time and from $O(n \log n)$ length trajectories of Glauber dynamics

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

- An Algorithmic Surprise
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- Taming Higher-Order Interactions

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

- An Algorithmic Surprise
- **The Trouble with Higher-Order Interactions**

Bonus Part: Population of Means

- Taming Higher-Order Interactions

First let's review **[Bresler, Gamarnik, Shah '15]**...

PROOF SKETCH (ISING CASE)

How can we decide if (X_u, X_v) is an edge?

PROOF SKETCH (ISING CASE)

How can we decide if (X_u, X_v) is an edge?

Definition: A **good sequence** is a contiguous sequence where

- (1) X_u attempts to update (might or might not flip)
- (2) Then X_v attempts to update, and flips
- (3) Then X_u attempts to update (might or might not flip)

PROOF SKETCH (ISING CASE)

How can we decide if (X_u, X_v) is an edge?

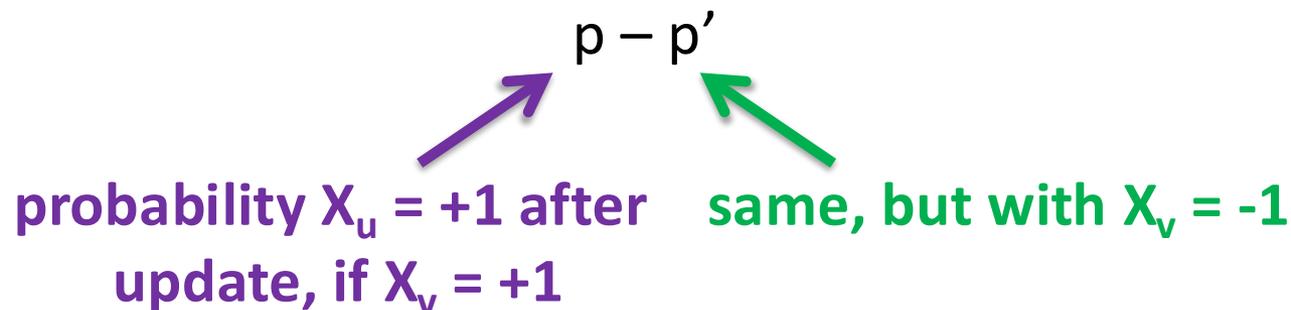
Definition: A **good sequence** is a contiguous sequence where

- (1) X_u attempts to update (might or might not flip)
- (2) Then X_v attempts to update, and flips
- (3) Then X_u attempts to update (might or might not flip)

Intuition: We get one sample from the conditional distribution on X_u when $X_v = +1$ and one sample when $X_v = -1$ but everything else stays the same

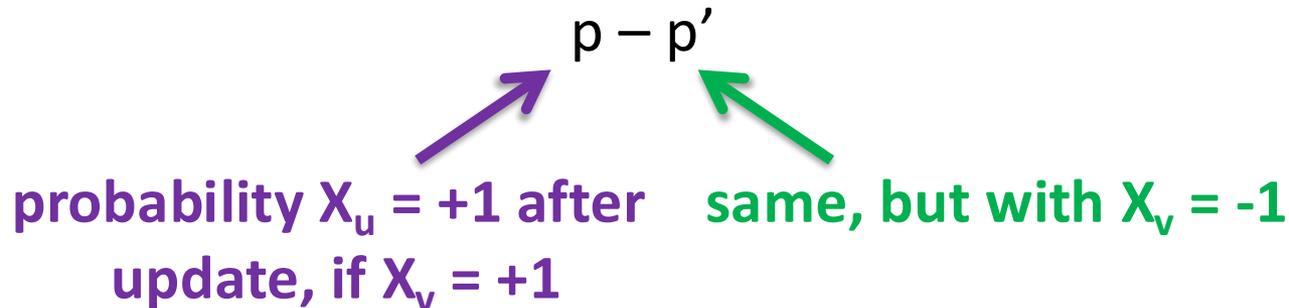
PROOF SKETCH (ISING CASE)

Good sequences give us empirical estimates of the quantity



PROOF SKETCH (ISING CASE)

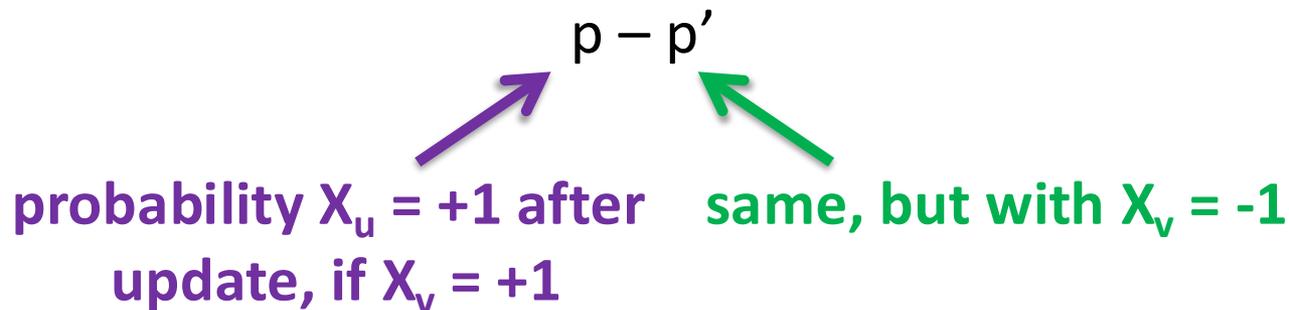
Good sequences give us empirical estimates of the quantity



The probabilities p and p' depend on the rest of the configuration, but have a consistent sign depending on $J_{u,v}$

PROOF SKETCH (ISING CASE)

Good sequences give us empirical estimates of the quantity

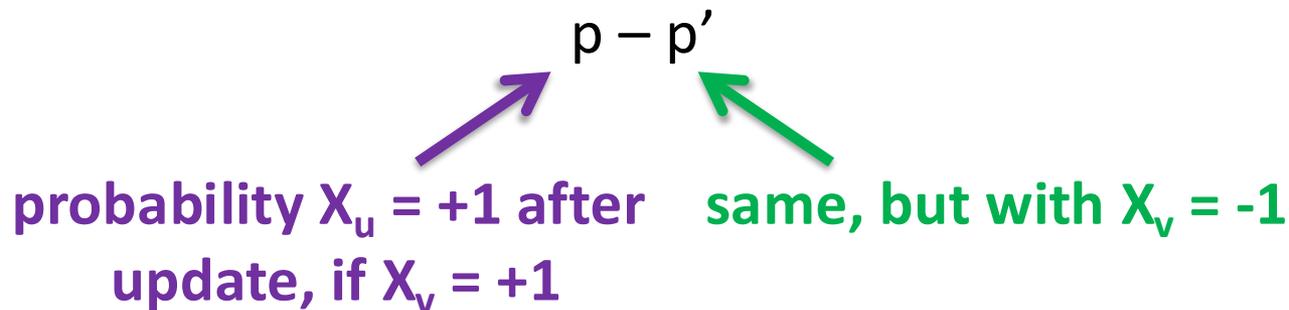


The probabilities p and p' depend on the rest of the configuration, but have a consistent sign depending on $J_{u,v}$

Claim: With high probability, edges correspond to pairs where the statistic averaged over good sequences is far enough from zero

PROOF SKETCH (ISING CASE)

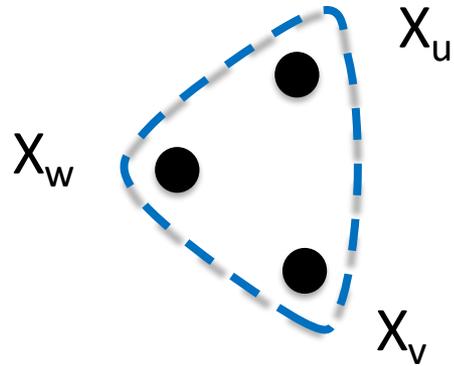
Good sequences give us empirical estimates of the quantity



 **Caution:** There is a minor bug because conditioning on X_v flipping changes things, can be fixed

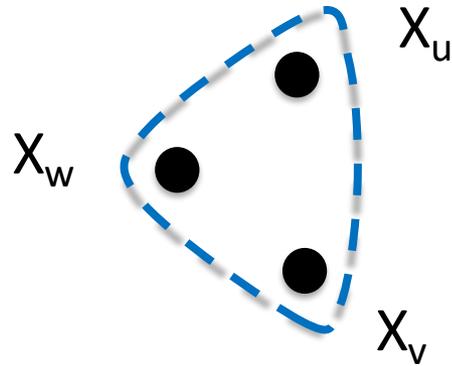
A STUMBLING BLOCK

Why does this strategy not work with higher-order interactions?



A STUMBLING BLOCK

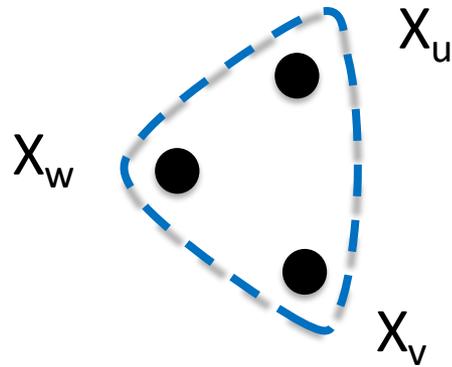
Why does this strategy not work with higher-order interactions?



Sometimes X_u and X_v want to agree, sometimes they want to disagree, depending on the sign of X_w

A STUMBLING BLOCK

Why does this strategy not work with higher-order interactions?



Sometimes X_u and X_v want to agree, sometimes they want to disagree, depending on the sign of X_w

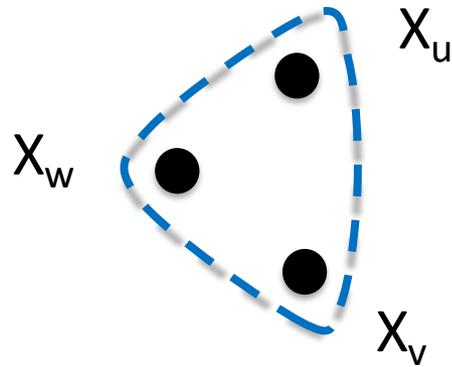
But now the quantity $p - p' \dots$

probability $X_u = +1$ after
update, if $X_v = +1$

same, but with $X_v = -1$

A STUMBLING BLOCK

Why does this strategy not work with higher-order interactions?



Sometimes X_u and X_v want to agree, sometimes they want to disagree, depending on the sign of X_w

But now the quantity $p - p'$

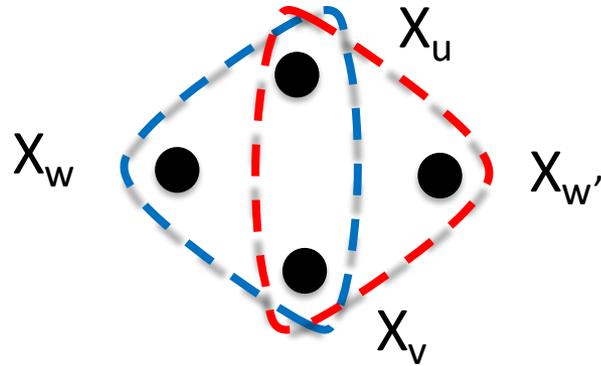
probability $X_u = +1$ after
update, if $X_v = +1$

same, but with $X_v = -1$

doesn't necessarily have a consistent sign

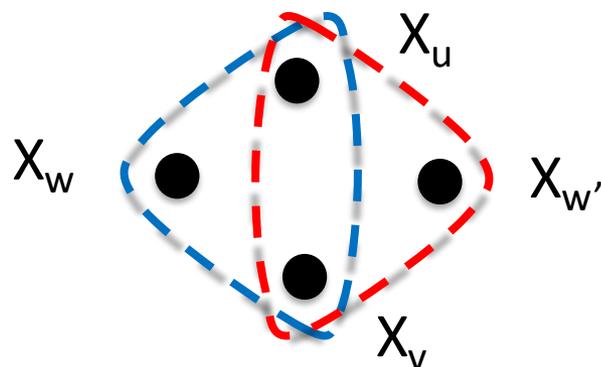
A STUMBLING BLOCK

Could also have situations like



A STUMBLING BLOCK

Could also have situations like



Where X_v affects the distribution of X_u through two paths, but they cancel out

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

- An Algorithmic Surprise
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- Taming Higher-Order Interactions

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

- An Algorithmic Surprise
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- Taming Higher-Order Interactions

POPULATION OF MEANS

Classic problem in statistics:

Setup: Suppose $\theta \sim \mu$ and $X_1, \dots, X_k \sim \text{Bern}(\theta)$

random student

their score on k exam questions

POPULATION OF MEANS

Classic problem in statistics:

Setup: Suppose $\theta \sim \mu$ and $X_1, \dots, X_k \sim \text{Bern}(\theta)$

random student

their score on k exam questions

Are students just randomly guessing?

POPULATION OF MEANS

Classic problem in statistics:

Setup: Suppose $\theta \sim \mu$ and $X_1, \dots, X_k \sim \text{Bern}(\theta)$

random student

their score on k exam questions

Are students just randomly guessing?

There's no hope of answering this question if $k=1$, but...

POPULATION OF MEANS

Classic problem in statistics:

Setup: Suppose $\theta \sim \mu$ and $X_1, \dots, X_k \sim \text{Bern}(\theta)$

random student

their score on k exam questions

Fact (shrinkage): We can't estimate the individual θ 's, but can estimate the first k moments of μ from batches

POPULATION OF MEANS

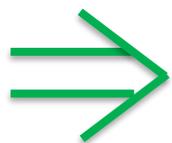
Classic problem in statistics:

Setup: Suppose $\theta \sim \mu$ and $X_1, \dots, X_k \sim \text{Bern}(\theta)$

random student

their score on k exam questions

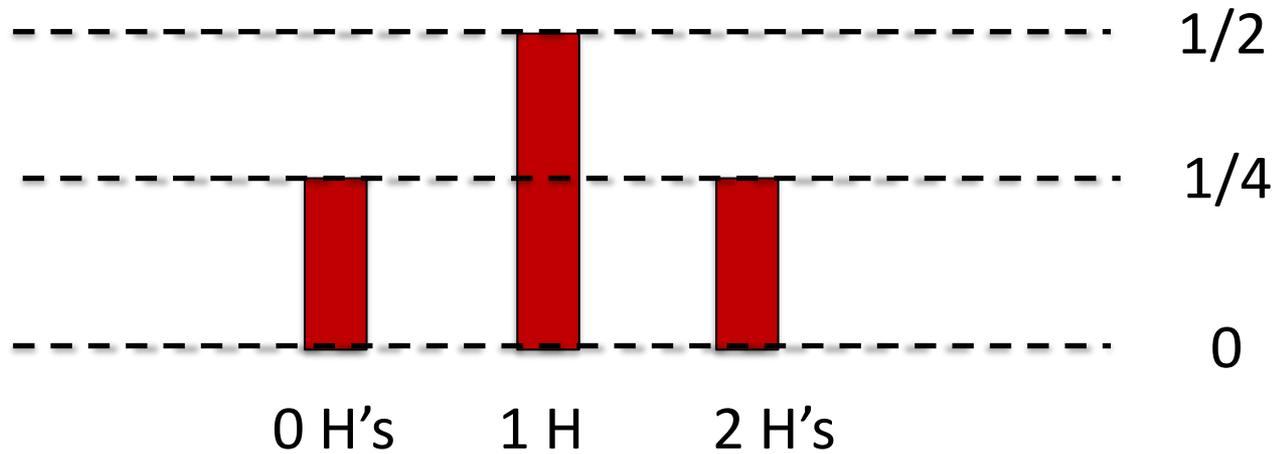
Fact (shrinkage): We can't estimate the individual θ 's, but can estimate the first k moments of μ from batches



Can decide if μ is a point mass even when $k = 2$

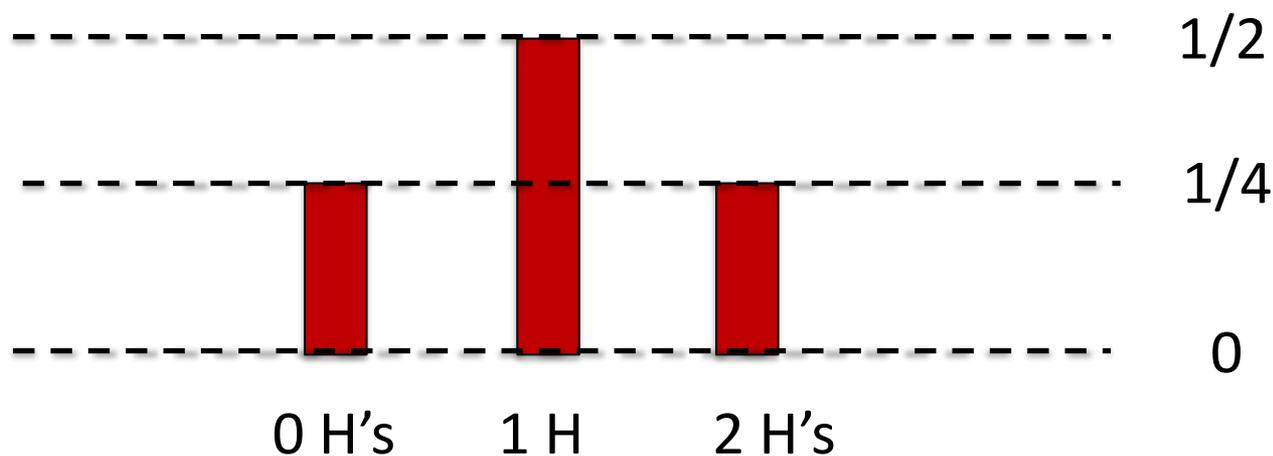
POPULATION OF MEANS

When $k = 2$ we can't estimate each student's θ well, but the point is the only way we can get



POPULATION OF MEANS

When $k = 2$ we can't estimate each student's θ well, but the point is the only way we can get



is if μ is a point mass on $1/2$, i.e. each student is randomly guessing

AN INTERESTING ASIDE

I learned this from Greg Valiant...

> [N Engl J Med.](#) 1979 Jun 28;300(26):1445-8. doi: 10.1056/NEJM197906283002601.

Gender of infants conceived on different days of the menstrual cycle

[S Harlap](#)

PMID: 449885 DOI: [10.1056/NEJM197906283002601](#)

AN INTERESTING ASIDE

I learned this from Greg Valiant...

> [N Engl J Med](#). 1979 Jun 28;300(26):1445-8. doi: 10.1056/NEJM197906283002601.

Gender of infants conceived on different days of the menstrual cycle

[S Harlap](#)

PMID: 449885 DOI: [10.1056/NEJM197906283002601](#)

...can test with population of means, from dizygotic twins...

AN INTERESTING ASIDE

I learned this from Greg Valiant...

> [N Engl J Med.](#) 1979 Jun 28;300(26):1445-8. doi: 10.1056/NEJM197906283002601.

Gender of infants conceived on different days of the menstrual cycle

[S Harlap](#)

PMID: 449885 DOI: [10.1056/NEJM197906283002601](#)

...can test with population of means, from dizygotic twins...
with data from dog kennels...

AN INTERESTING ASIDE

I learned this from Greg Valiant...

> [N Engl J Med.](#) 1979 Jun 28;300(26):1445-8. doi: 10.1056/NEJM197906283002601.

Gender of infants conceived on different days of the menstrual cycle

[S Harlap](#)

PMID: 449885 DOI: [10.1056/NEJM197906283002601](#)

...can test with population of means, from dizygotic twins...
with data from dog kennels...but no evidence found

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

- An Algorithmic Surprise
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- Taming Higher-Order Interactions

OUTLINE

Part I: Introduction and Review

- Computational Lower Bounds

Part II: Learning from Trajectories

- An Algorithmic Surprise
- The Trouble with Higher-Order Interactions

Bonus Part: Population of Means

- **Taming Higher-Order Interactions**

KEY IDEA

Can we instead estimate the following quantity?

$$(p - p')^2$$

probability $X_u = +1$ after
update, if $X_v = +1$

same, but with $X_v = -1$

KEY IDEA

Can we instead estimate the following quantity?

$$(p - p')^2$$

probability $X_u = +1$ after
update, if $X_v = +1$

same, but with $X_v = -1$

All we need is a more complicated definition of a good sequence

KEY IDEA

Definition': A **better sequence** is a contiguous sequence where

- (1) X_u attempts to update ^{twice} (might or might not flip)

- (2) Then X_v attempts to update, and flips
- (3) Then X_u attempts to update ^{twice} (might or might not flip)


KEY IDEA

Definition': A **better sequence** is a contiguous sequence where

- (1) X_u attempts to update ^{twice} (might or might not flip)

- (2) Then X_v attempts to update, and flips
- (3) Then X_u attempts to update ^{twice} (might or might not flip)


Intuition: We get **two** samples from the conditional distribution on X_u when $X_v = +1$ and **two** samples when $X_v = -1$ but everything else stays the same

KEY IDEA

Definition': A **better sequence** is a contiguous sequence where

- (1) X_u attempts to update ^{twice} (might or might not flip)

- (2) Then X_v attempts to update, and flips
- (3) Then X_u attempts to update ^{twice} (might or might not flip)


Let $Y_1 Y_2 \dots Y'_1 Y'_2$ be the indicators that $X_u = +1$ at corresponding locations

KEY IDEA

Definition': A **better sequence** is a contiguous sequence where

- (1) X_u attempts to update ^{twice} (might or might not flip)
- (2) Then X_v attempts to update, and flips
- (3) Then X_u attempts to update ^{twice} (might or might not flip)

Let $Y_1 Y_2 \dots Y'_1 Y'_2$ be the indicators that $X_u = +1$ at corresponding locations

Now we can estimate $(p-p')^2$ using the statistic

$$Z = Y_1 Y_2 - 2Y_1 Y'_1 + Y'_1 Y'_2$$

EPILOGUE

This strategy would need very long trajectories...

EPILOGUE

This strategy would need very long trajectories...

Idea (informal): Even if the attempted updates to X_u and X_v are not contiguous, if they're close enough we don't expect any updates to their neighbors in between

EPILOGUE

This strategy would need very long trajectories...

Idea (informal): Even if the attempted updates to X_u and X_v are not contiguous, if they're close enough we don't expect any updates to their neighbors in between

Also can learn from just the updates:

Theorem [Gaitonde, Moitra, Mossel '26]: Can learn Ising models from just the evolution of the system in continuous time

Summary:

- Learning Markov Random Fields from samples is computationally hard
- But you can't generate samples anyways and given the trajectories there are much better algorithms
- **What other learning problems become easier from dynamics?**

Summary:

- Learning Markov Random Fields from samples is computationally hard
- But you can't generate samples anyways and given the trajectories there are much better algorithms
- **What other learning problems become easier from dynamics?**

Thanks! Any Questions?