Computation, Statistics, and Optimization of Random Functions

Afonso S. Bandeira

ETH zürich

Mathematics, Physics and Machine Learning (IST, Lisbon),
June 4, 2020

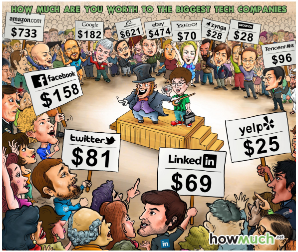
The Age of Data

- *"The world's most valuable resource is no longer oil, but data"*
  – The Economist

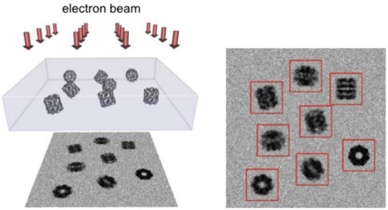
The Age of Data

- “*The world’s most valuable resource is no longer oil, but data*”
  – The Economist

- “*We estimate AI-powered applications will add $13 trillion in value to the global economy in the coming decade*”
  – McKinsey & Company



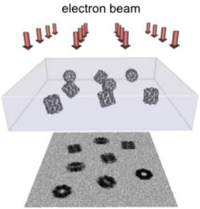
Cryo-Electron Microscopy

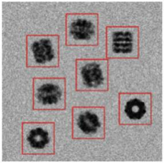


Task: Reconstruct the 3d molecule from noisy projections taken from unknown directions

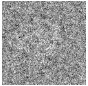
Images courtesy of Amit Singer, Yoel Shkolnisky, and Fred Sigworth

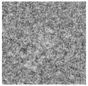
Cryo-Electron Microscopy





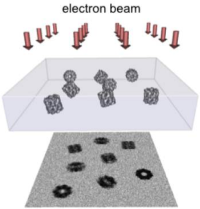
Task: Reconstruct the 3d molecule from noisy projections taken from unknown directions

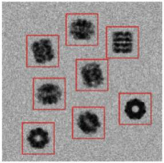




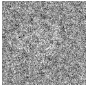
Images courtesy of Amit Singer, Yoel Shkolnisky, and Fred Sigworth

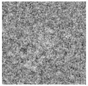
Cryo-Electron Microscopy

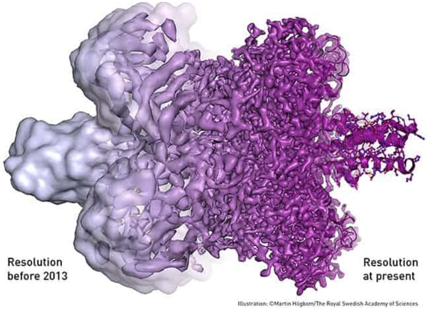




Task: Reconstruct the 3d molecule from noisy projections taken from unknown directions

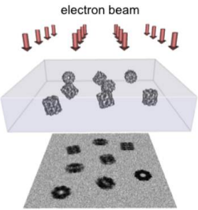


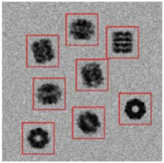




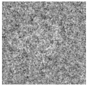
Images courtesy of Amit Singer, Yoel Shkolnisky, and Fred Sigworth

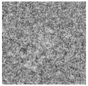
Cryo-Electron Microscopy

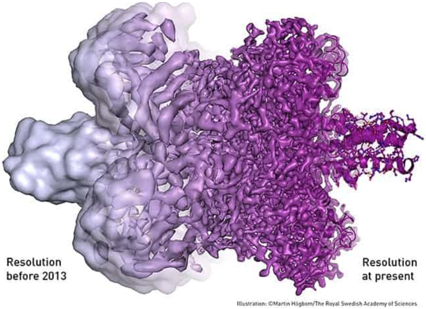




Task: Reconstruct the $3d$ molecule from noisy projections taken from unknown directions







2017 Chemistry Laureates. Ill: N. Elmehed.

2017 Nobel Prize in Chemistry

The Nobel Prize in Chemistry 2017 was awarded to Jacques Dubochet, Joachim Frank and Richard Henderson *"for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution"*.

Images courtesy of Amit Singer, Yoel Shkolnisky, and Fred Sigworth

Mathematics of Data

- Are there **limits** to what we can learn?

Mathematics of Data

- Are there **limits** to what we can learn?
- Which methods work?

  Why?

– XKCD

Mathematics of Data

- Are there **limits** to what we can learn?
- Which methods work? **Why?**
- What are the **bottlenecks**?

– XKCD

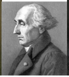
Mathematics of Data

- Are there **limits** to what we can learn?
- Which methods work? **Why?**
- What are the **bottlenecks**?
- Can we *a posteriori* **certify**?

– XKCD

Statistics — What are limits to learning?

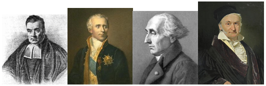
- 1700's - **Bayesian Statistics**

Bayes Laplace Lagrange Gauss
1760's 1770's 1800's

Statistics — What are limits to learning?

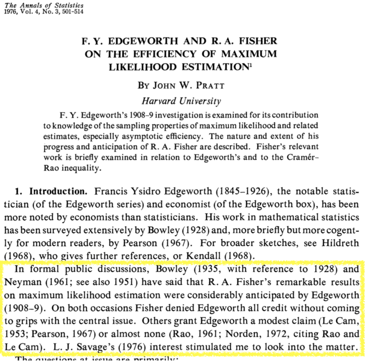
- 1700's - **Bayesian Statistics**
- 1900-1920 - **Fisher Information**
  — How much information about a parameter does a sample have?



Bayes 1760's Laplace Lagrange 1770's Gauss 1800's

Statistics — What are limits to learning?

- 1700's - **Bayesian Statistics**
- 1900-1920 - **Fisher Information** — How much information about a parameter does a sample have?


The Annals of Statistics
1976, Vol. 4, No. 3, 501–514

F. Y. EDGEWORTH AND R. A. FISHER ON THE EFFICIENCY OF MAXIMUM LIKELIHOOD ESTIMATION[1]

BY JOHN W. PRATT

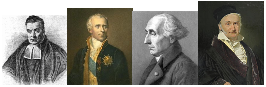
Harvard University

F. Y. Edgeworth's 1908–9 investigation is examined for its contribution to knowledge of the sampling properties of maximum likelihood and related estimates, especially asymptotic efficiency. The nature and extent of his progress and anticipation of R. A. Fisher are described. Fisher's relevant work is briefly examined in relation to Edgeworth's and to the Cramér–Rao inequality.

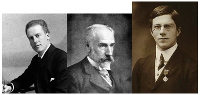
1. Introduction. Francis Ysidro Edgeworth (1845–1926), the notable statistician (of the Edgeworth series) and economist (of the Edgeworth box), has been more noted by economists than statisticians. His work in mathematical statistics has been surveyed extensively by Bowley (1928) and, more briefly but more cogently for modern readers, by Pearson (1967). For broader sketches, see Hildreth (1968), who gives further references, or Kendall (1968).

In formal public discussions, Bowley (1935, with reference to 1928) and Neyman (1961; see also 1951) have said that R. A. Fisher's remarkable results on maximum likelihood estimation were considerably anticipated by Edgeworth (1908–9). On both occasions Fisher denied Edgeworth all credit without coming to grips with the central issue. Others grant Edgeworth a modest claim (Le Cam, 1953; Pearson, 1967) or almost none (Rao, 1961; Norden, 1972, citing Rao and Le Cam). L. J. Savage's (1976) interest stimulated me to look into the matter.

The questions at issue are primarily:



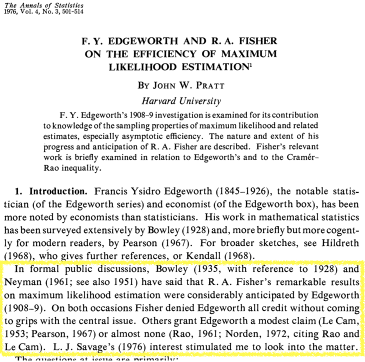
Bayes 1760's | Laplace | Lagrange 1770's | Gauss 1800's



K.Pearson 1890's | Edgeworth 1900's | Fisher 1920's

Statistics — What are limits to learning?

- 1700's - **Bayesian Statistics**
- 1900-1920 - **Fisher Information**
  — How much information about a parameter does a sample have?
- 1933: **Neyman-Pearson Lemma**:
  — Limits on Hypothesis Testing



The Annals of Statistics
1976, Vol. 4, No. 3, 501-514

F. Y. EDGEWORTH AND R. A. FISHER ON THE EFFICIENCY OF MAXIMUM LIKELIHOOD ESTIMATION[1]

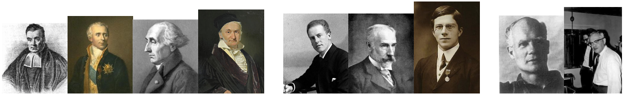
BY JOHN W. PRATT

Harvard University

F. Y. Edgeworth's 1908–9 investigation is examined for its contribution to knowledge of the sampling properties of maximum likelihood and related estimates, especially asymptotic efficiency. The nature and extent of his progress and anticipation of R. A. Fisher are described. Fisher's relevant work is briefly examined in relation to Edgeworth's and to the Cramér-Rao inequality.

1. Introduction. Francis Ysidro Edgeworth (1845–1926), the notable statistician (of the Edgeworth series) and economist (of the Edgeworth box), has been more noted by economists than statisticians. His work in mathematical statistics has been surveyed extensively by Bowley (1928) and, more briefly but more cogently for modern readers, by Pearson (1967). For broader sketches, see Hildreth (1968), who gives further references, or Kendall (1968).

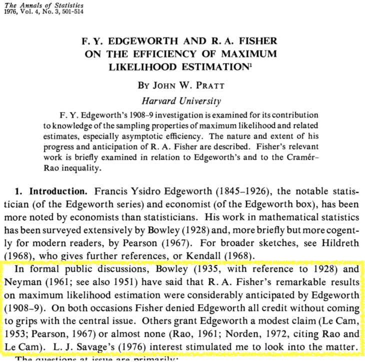
In formal public discussions, Bowley (1935, with reference to 1928) and Neyman (1961; see also 1951) have said that R. A. Fisher's remarkable results on maximum likelihood estimation were considerably anticipated by Edgeworth (1908–9). On both occasions Fisher denied Edgeworth all credit without coming to grips with the central issue. Others grant Edgeworth a modest claim (Le Cam, 1953; Pearson, 1967) or almost none (Rao, 1961; Norden, 1972, citing Rao and Le Cam). L. J. Savage's (1976) interest stimulated me to look into the matter.



Bayes 1760's | Laplace | Lagrange 1770's | Gauss 1800's | K.Pearson 1890's | Edgeworth 1900's | Fisher 1920's | E.Pearson | Neyman 1930's

Statistics — What are limits to learning?

- 1700's - **Bayesian Statistics**
- 1900-1920 - **Fisher Information**
  — How much information about a parameter does a sample have?
- 1933: **Neyman-Pearson Lemma**:
  — Limits on Hypothesis Testing
- 1940's: **Cramér-Rao Bound**:
  — Limits on Statistical Estimation



The Annals of Statistics
1976, Vol. 4, No. 3, 501–514

F. Y. EDGEWORTH AND R. A. FISHER ON THE EFFICIENCY OF MAXIMUM LIKELIHOOD ESTIMATION[1]

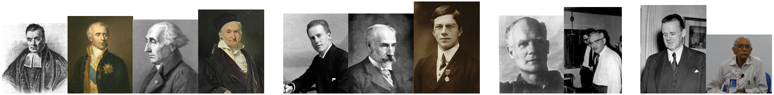
BY JOHN W. PRATT

Harvard University

F. Y. Edgeworth's 1908–9 investigation is examined for its contribution to knowledge of the sampling properties of maximum likelihood and related estimates, especially asymptotic efficiency. The nature and extent of his progress and anticipation of R. A. Fisher are described. Fisher's relevant work is briefly examined in relation to Edgeworth's and to the Cramér–Rao inequality.

1. Introduction. Francis Ysidro Edgeworth (1845–1926), the notable statistician (of the Edgeworth series) and economist (of the Edgeworth box), has been more noted by economists than statisticians. His work in mathematical statistics has been surveyed extensively by Bowley (1928) and, more briefly but more cogently for modern readers, by Pearson (1967). For broader sketches, see Hildreth (1968), who gives further references, or Kendall (1968).

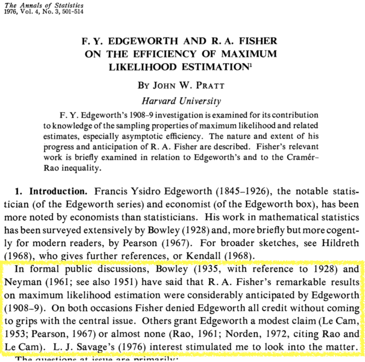
In formal public discussions, Bowley (1935, with reference to 1928) and Neyman (1961; see also 1951) have said that R. A. Fisher's remarkable results on maximum likelihood estimation were considerably anticipated by Edgeworth (1908–9). On both occasions Fisher denied Edgeworth all credit without coming to grips with the central issue. Others grant Edgeworth a modest claim (Le Cam, 1953; Pearson, 1967) or almost none (Rao, 1961; Norden, 1972, citing Rao and Le Cam). L. J. Savage's (1976) interest stimulated me to look into the matter.



Bayes 1760's | Laplace | Lagrange 1770's | Gauss 1800's | K.Pearson 1890's | Edgeworth 1900's | Fisher 1920's | E.Pearson | Neyman 1930's | Cramér | Rao 1940's

Statistics — What are limits to learning?

- 1700's - **Bayesian Statistics**
- 1900-1920 - **Fisher Information**
  — How much information about a parameter does a sample have?
- 1933: **Neyman-Pearson Lemma**:
  — Limits on Hypothesis Testing
- 1940's: **Cramér-Rao Bound**:
  — Limits on Statistical Estimation
- 1950+ Minimax, Contiguity, ...



The Annals of Statistics
1976, Vol. 4, No. 3, 501–514

F. Y. EDGEWORTH AND R. A. FISHER ON THE EFFICIENCY OF MAXIMUM LIKELIHOOD ESTIMATION[1]

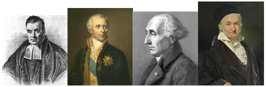
BY JOHN W. PRATT

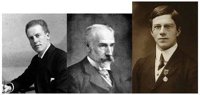
Harvard University

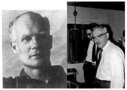
F. Y. Edgeworth's 1908–9 investigation is examined for its contribution to knowledge of the sampling properties of maximum likelihood and related estimates, especially asymptotic efficiency. The nature and extent of his progress and anticipation of R. A. Fisher are described. Fisher's relevant work is briefly examined in relation to Edgeworth's and to the Cramér–Rao inequality.

1. Introduction. Francis Ysidro Edgeworth (1845–1926), the notable statistician (of the Edgeworth series) and economist (of the Edgeworth box), has been more noted by economists than statisticians. His work in mathematical statistics has been surveyed extensively by Bowley (1928) and, more briefly but more cogently for modern readers, by Pearson (1967). For broader sketches, see Hildreth (1968), who gives further references, or Kendall (1968).

In formal public discussions, Bowley (1935, with reference to 1928) and Neyman (1961; see also 1951) have said that R. A. Fisher's remarkable results on maximum likelihood estimation were considerably anticipated by Edgeworth (1908–9). On both occasions Fisher denied Edgeworth all credit without coming to grips with the central issue. Others grant Edgeworth a modest claim (Le Cam, 1953; Pearson, 1967) or almost none (Rao, 1961; Norden, 1972, citing Rao and Le Cam). L. J. Savage's (1976) interest stimulated me to look into the matter.

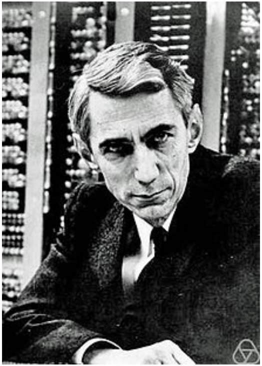






Bayes 1760's | Laplace | Lagrange 1770's | Gauss 1800's | K.Pearson 1890's | Edgeworth 1900's | Fisher 1920's | E.Pearson | Neyman 1930's | Cramér | Rao 1940's | ...

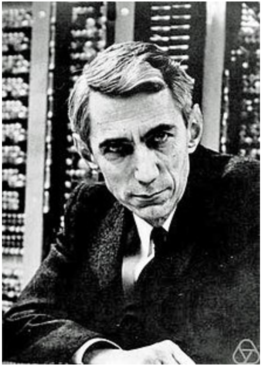
Information Theory



Claude Shanon '48:
A Mathematical Theory of Communication
Shannon Entropy

Richard Hamming '50:
"Error detecting and error correcting codes"
Hamming Distance

Information Theory



Claude Shanon '48:
A Mathematical Theory of Communication
Shannon Entropy

Richard Hamming '50:
"Error detecting and error correcting codes"
Hamming Distance

Shannon Entropy: # of bits "of information" needed to identify a draw of X

Learning/Estimating is (also) optimization

Goal: Find parameter/signal/model that best "fits" the data

- Maximum likelihood estimation
- Training of Neural Networks
- ...

Learning/Estimating is (also) optimization

Goal: Find parameter/signal/model that best "fits" the data

- Maximum likelihood estimation
- Training of Neural Networks
- ...

Are these computational tasks feasible/easy?

Learning/Estimating is (also) optimization

Goal: Find parameter/signal/model that best "fits" the data

- Maximum likelihood estimation
- Training of Neural Networks
- ...

Are these computational tasks feasible/easy?

Many optimization/computational problems are **NP-hard** (e.g. Knapsack)

Learning/Estimating is (also) optimization

Goal: Find parameter/signal/model that best "fits" the data

- Maximum likelihood estimation
- Training of Neural Networks
- . . .

Are these computational tasks feasible/easy?

Many optimization/computational problems are **NP-hard** (e.g. Knapsack)

1956: Gödel's letter to von Neumann (and John Nash's 1955)

Learning/Estimating is (also) optimization

Goal: Find parameter/signal/model that best "fits" the data

- Maximum likelihood estimation
- Training of Neural Networks
- ...

Are these computational tasks feasible/easy?

Many optimization/computational problems are **NP-hard** (e.g. Knapsack)

1956: Gödel's letter to von Neumann (and John Nash's 1955)

1971-72: Cook and Karp's **NP-hardness**

Learning/Estimating is (also) optimization

Goal: Find parameter/signal/model that best "fits" the data

- Maximum likelihood estimation
- Training of Neural Networks
- ...

Are these computational tasks feasible/easy?

Many optimization/computational problems are **NP-hard** (e.g. Knapsack)

1956: Gödel's letter to von Neumann (and John Nash's 1955)

1971-72: Cook and Karp's **NP-hardness**

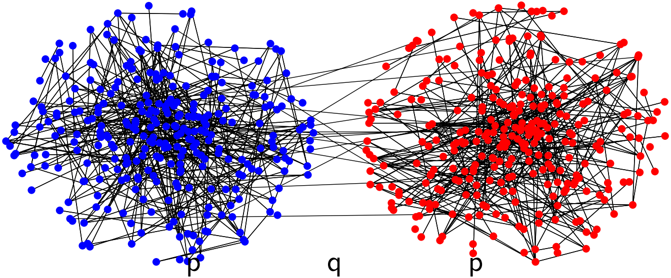
Should we design (statistical) models so that optimization is easy?

Linearity, Convexity, ...

An example: Communities in Social Networks

Given two disjoint sets of $m = \frac{n}{2}$ nodes each. Independently:

- pairs between clusters have an edge with probability p
- pairs across clusters have an edge with probability $q < p$



A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, 2011

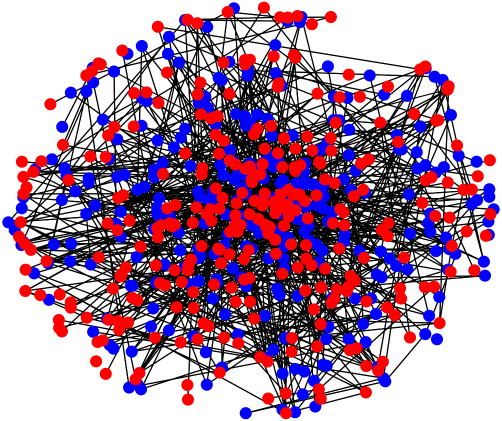
E. Mossel. J. Neeman, A. Sly, 2012, 2013.

L. Massoulie, 2013.

E. Abbe, A. S. Bandeira, G. Hall, 2014.

A. S. Bandeira, 2018.

An example: Communities in Social Networks



A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, 2011

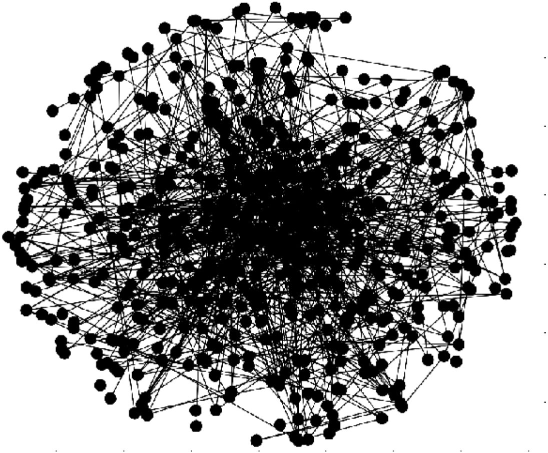
E. Mossel. J. Neeman, A. Sly, 2012, 2013.

L. Massoulie, 2013.

E. Abbe, A. S. Bandeira, G. Hall, 2014.

A. S. Bandeira, 2018.

An example: Communities in Social Networks



Can we recover the labels?

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, 2011

E. Mossel. J. Neeman, A. Sly, 2012, 2013.

L. Massoulie, 2013.

E. Abbe, A. S. Bandeira, G. Hall, 2014.

A. S. Bandeira, 2018.

An example: continued

- **Theorem:** For $p = \alpha \frac{\log n}{n}$ and $q = \beta \frac{\log n}{n}$, If (iff)

$$\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2},$$

the **Minimum Bisection** coincides with the true communities.

E. Abbe, A. S. Bandeira, G. Hall, 2014.

E. Mossel, J. Neeman, and A. Sly, 2014

B. Hajek, Y. Wu, and J. Xu., 2014

A. S. Bandeira, 2015.

An example: continued

- **Theorem:** For $p = \alpha \frac{\log n}{n}$ and $q = \beta \frac{\log n}{n}$, If (iff)

$$\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2},$$

  the **Minimum Bisection** coincides with the true communities.

- **Theorem:** **Minimum Bisection is an NP-hard problem.**

E. Abbe, A. S. Bandeira, G. Hall, 2014.

E. Mossel, J. Neeman, and A. Sly, 2014

B. Hajek, Y. Wu, and J. Xu., 2014

A. S. Bandeira, 2015.

An example: continued

- **Theorem:** For $p = \alpha \frac{\log n}{n}$ and $q = \beta \frac{\log n}{n}$, If (iff)

$$\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2},$$

  the **Minimum Bisection** coincides with the true communities.

- **Theorem:** **Minimum Bisection is an NP-hard problem.**

- **Theorem:** If

$$\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2},$$

  Minimum Bisection can be computed efficiently.

E. Abbe, A. S. Bandeira, G. Hall, 2014.

E. Mossel, J. Neeman, and A. Sly, 2014

B. Hajek, Y. Wu, and J. Xu., 2014

A. S. Bandeira, 2015.

An example: continued

- **Theorem:** For $p = \alpha\frac{\log n}{n}$ and $q = \beta\frac{\log n}{n}$, If (iff)

$$\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2},$$

the **Minimum Bisection** coincides with the true communities.

- **Theorem: Minimum Bisection is an NP-hard problem.**
- **Theorem:** If

$$\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2},$$

Minimum Bisection can be computed efficiently.

Does this always happen?

E. Abbe, A. S. Bandeira, G. Hall, 2014.

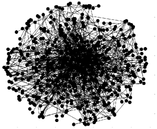
E. Mossel, J. Neeman, and A. Sly, 2014

B. Hajek, Y. Wu, and J. Xu., 2014

A. S. Bandeira, 2015.

Statistical-to-Computational Gaps

Hidden Clique Problem

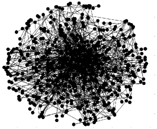


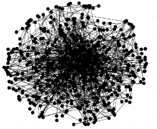
- A graph $G\left(n, \frac{1}{2}\right)$

  — each edge appears with probability $\frac{1}{2}$

Statistical-to-Computational Gaps

Hidden Clique Problem





- A graph $G\left(n, \frac{1}{2}\right)$

  — each edge appears with probability $\frac{1}{2}$

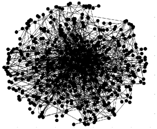
Vs

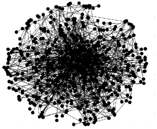
- $G\left(n, \frac{1}{2}\right) + \mathbf{k}$-clique

  k picked at random and all the edges between them added

Statistical-to-Computational Gaps

Hidden Clique Problem





- A graph $G\left(n, \frac{1}{2}\right)$

  — each edge appears with probability $\frac{1}{2}$

Vs

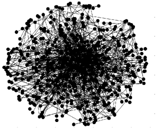
- $G\left(n, \frac{1}{2}\right) + \mathbf{k}$-clique

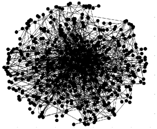
  k picked at random and all the edges between them added

$2 \log n$	**Largest Clique**	**k**

Statistical-to-Computational Gaps

Hidden Clique Problem





- A graph $G\left(n, \frac{1}{2}\right)$

  — each edge appears with probability $\frac{1}{2}$

Vs

- $G\left(n, \frac{1}{2}\right) + \mathbf{k}$-clique

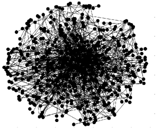
  $\mathbf{k}$ picked at random and all the edges between them added

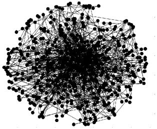
$2 \log n$ — **Largest Clique** — $\mathbf{k}$

- *Alon-Krivelevich-Sudakov '98:* Efficient algorithm for $\mathbf{k} \gtrsim \sqrt{n}$ (as opposed to $\mathbf{k} > 2 \log n$)

Statistical-to-Computational Gaps

Hidden Clique Problem





- A graph $G\left(n, \frac{1}{2}\right)$

  — each edge appears with probability $\frac{1}{2}$

Vs

- $G\left(n, \frac{1}{2}\right) + \mathbf{k}$-clique

  $\mathbf{k}$ picked at random and all the edges between them added

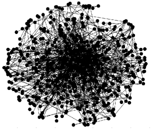
Largest Clique

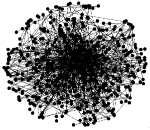
$2 \log n$ $\mathbf{k}$

- *Alon-Krivelevich-Sudakov '98:* Efficient algorithm for $\mathbf{k} \gtrsim \sqrt{n}$ (as opposed to $\mathbf{k} > 2 \log n$)
- No improvement since; believed to be **hard** and used as reduction primitive (e.g. Berthet-Rigollet '12)

Statistical-to-Computational Gaps

Hidden Clique Problem





- A graph $G\left(n, \frac{1}{2}\right)$

  — each edge appears with probability $\frac{1}{2}$

Vs

- $G\left(n, \frac{1}{2}\right) + \mathbf{k}$-clique

  k picked at random and all the edges between them added

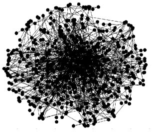
$2 \log n$ **Largest Clique** **k**

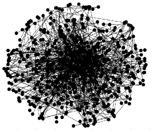
- *Alon-Krivelevich-Sudakov '98:* Efficient algorithm for $\mathbf{k} \gtrsim \sqrt{n}$ (as opposed to $\mathbf{k} > 2 \log n$)
- No improvement since; believed to be **hard** and used as reduction primitive (e.g. Berthet-Rigollet '12)

Statistical-to-Computational Gap "Hypothesis"

Statistical-to-Computational Gaps

Hidden Clique Problem





- A graph $G\left(n, \frac{1}{2}\right)$

  — each edge appears with probability $\frac{1}{2}$

Vs

- $G\left(n, \frac{1}{2}\right) + \mathbf{k}$-clique

  k picked at random and all the edges between them added

	Largest Clique	
$2 \log n$		$\mathbf{k}$

- *Alon-Krivelevich-Sudakov '98:* Efficient algorithm for $\mathbf{k} \gtrsim \sqrt{n}$ (as opposed to $\mathbf{k} > 2 \log n$)
- No improvement since; believed to be **hard** and used as reduction primitive (e.g. Berthet-Rigollet '12)

Statistical-to-Computational Gap "Hypothesis"

What Makes a Problem Hard?

Complexity/Geometry of **Posterior/Solutions**

$\mathbb{P}(\text{ node labels } | \text{ SBM Graph })$ $\leftrightarrow$ Spin Glass (Physics)

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, 2011

D. Gamarnik, M. Sudan, 2013

A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, 2018

What Makes a Problem Hard?

Complexity/Geometry of **Posterior/Solutions**

$\mathbb{P}(\text{ node labels} \mid \text{SBM Graph })$ $\leftrightarrow$ Spin Glass (Physics)

What is the geometry of cliques $\geq \omega$ in $G\left(n, \frac{1}{2}\right)$

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, 2011

D. Gamarnik, M. Sudan, 2013

A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, 2018

What Makes a Problem Hard?

Complexity/Geometry of **Posterior/Solutions**

$\mathbb{P}(\text{ node labels} \mid \text{SBM Graph })$ $\leftrightarrow$ Spin Glass (Physics)

What is the geometry of cliques $\geq \omega$ in $G\left(n, \frac{1}{2}\right)$

- Community Detection in the Stochastic Block Model has a **computational gap** for ≥ 5 communities

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, 2011

D. Gamarnik, M. Sudan, 2013

A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, 2018

What Makes a Problem Hard?

Complexity/Geometry of **Posterior/Solutions**

$\mathbb{P}$(node labels | SBM Graph) $\leftrightarrow$ Spin Glass (Physics)

What is the geometry of cliques $\geq \omega$ in $G\left(n, \frac{1}{2}\right)$

- Community Detection in the Stochastic Block Model has a **computational gap** for ≥ 5 communities
- Finding a clique of size $(1+\varepsilon)\log n$ in a $G\left(n, \frac{1}{2}\right)$ is **hard**

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, 2011

D. Gamarnik, M. Sudan, 2013

A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, 2018

What Makes a Problem Hard?

Complexity/Geometry of **Posterior/Solutions**

$\mathbb{P}$(node labels | SBM Graph) $\leftrightarrow$ Spin Glass (Physics)

What is the geometry of cliques $\geq \omega$ in $G\left(n, \frac{1}{2}\right)$

- Community Detection in the Stochastic Block Model has a **computational gap** for ≥ 5 communities
- Finding a clique of size $(1+\varepsilon)\log n$ in a $G\left(n, \frac{1}{2}\right)$ is **hard**
- Many versions of structured Random Matrix Spike Models have a **computational gap** in recovery
- ...

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, 2011

D. Gamarnik, M. Sudan, 2013

A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, 2018

Algebraic Considerations

Sum-of-Square: A Hierarchy of algorithms
inspired on **Hilbert Nullstellensatz** (Parrilo '00, Lassere '01, . . .)

Y. Ding, D. Kunisky, A. S. Wein, A. S. Bandeira, 2019

A. S. Bandeira, D. Kunisky, A. S. Wein, 2020

Algebraic Considerations

Sum-of-Square: A Hierarchy of algorithms
inspired on **Hilbert Nullstellensatz** (Parrilo '00, Lassere '01, . . .)

What if we restrict to **low-degree polynomials** of the data?

(Hopkins-Steurer '17, . . .)

Y. Ding, D. Kunisky, A. S. Wein, A. S. Bandeira, 2019

A. S. Bandeira, D. Kunisky, A. S. Wein, 2020

Algebraic Considerations

Sum-of-Square: A Hierarchy of algorithms
inspired on **Hilbert Nullstellensatz** (Parrilo '00, Lassere '01, ...)

What if we restrict to **low-degree polynomials** of the data?
(Hopkins-Steurer '17, ...)

- Exploiting sparsity ρn in **Sparse PCA** requires $\exp(\rho^2 \mathbf{n})$ **computation**
  $x_k \sim \mathcal{N}(0, I + \beta x x^T), \quad \|x\|_0 = \rho n$

Y. Ding, D. Kunisky, A. S. Wein, A. S. Bandeira, 2019

A. S. Bandeira, D. Kunisky, A. S. Wein, 2020

Algebraic Considerations

Sum-of-Square: A Hierarchy of algorithms
inspired on **Hilbert Nullstellensatz** (Parrilo '00, Lassere '01, . . .)

What if we restrict to **low-degree polynomials** of the data?
(Hopkins-Steurer '17, . . .)

- Exploiting sparsity ρn in **Sparse PCA** requires $\exp(\rho^2 \mathbf{n})$ **computation**
  $$x_k \sim \mathcal{N}(0, I + \beta x x^T), \quad \|x\|_0 = \rho n$$
- **Certifying** a non-trivial upper bound on the **Sherrington-Kirkpatrick Hamiltonian** is **hard**
  $$\max_{x \in \{\pm 1\}^n} x^T W x$$
  $$W_{ij} \sim \mathcal{N}(0, 1)$$
- . . .

Y. Ding, D. Kunisky, A. S. Wein, A. S. Bandeira, 2019

A. S. Bandeira, D. Kunisky, A. S. Wein, 2020

Statistics and Computation in Cryo-EM

- Connection between **Statistics of Cryo-EM** and **Algebraic Invariant Theory** gives:

$$\textbf{Optimal Reconstruction Quality} \sim \sqrt{\#\text{ of samples}} \times \mathrm{SNR}^{\mathbf{3}}$$

Bandeira, Niles-Weed, Rigollet, 2017.

Perry, Weed, Bandeira, Rigollet, Singer, 2017.

Bandeira, Blum-Smith, Kileel, Perry, Weed, Wein, 2017.

Statistics and Computation in Cryo-EM

- Connection between **Statistics of Cryo-EM**
  and **Algebraic Invariant Theory** gives:

$$\textbf{Optimal Reconstruction Quality} \sim \sqrt{\#\text{ of samples}} \times \text{SNR}^{3}$$

No computational gap!

Bandeira, Niles-Weed, Rigollet, 2017.

Perry, Weed, Bandeira, Rigollet, Singer, 2017.

Bandeira, Blum-Smith, Kileel, Perry, Weed, Wein, 2017.

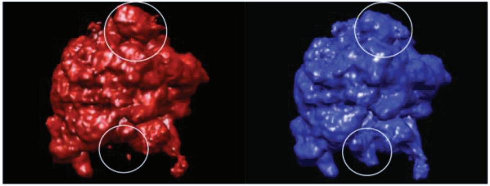
Statistics and Computation in Cryo-EM

- Connection between **Statistics of Cryo-EM** and **Algebraic Invariant Theory** gives:

$$\textbf{Optimal Reconstruction Quality} \sim \sqrt{\#\text{ of samples}} \times \mathrm{SNR}^{3}$$

No computational gap!

- **Computational gap** believed to arise in **Heterogeneity problem**

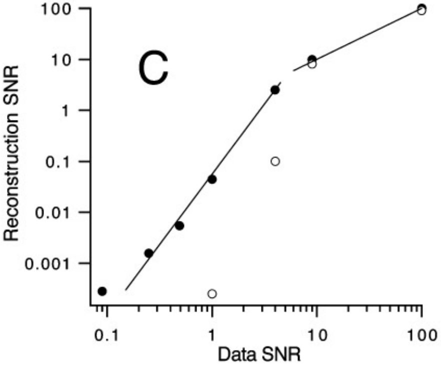


Bandeira, Niles-Weed, Rigollet, 2017.

Perry, Weed, Bandeira, Rigollet, Singer, 2017.

Bandeira, Blum-Smith, Kileel, Perry, Weed, Wein, 2017.

Behavior observed 20 years ago!



- The surprising $1/\mathrm{SNR}^3$ scaling at low SNR was observed in '98

F. Sigworth, Journal of Structural Biology, 1998.

Other Methods

- **Reductions** — If X is hard, so is Y

Other Methods

- **Reductions** — If X is hard, so is Y
- Geometry of Random Optimization Landscapes

  e.g.: **Kac-Rice formula**

Other Methods

- **Reductions** — If X is hard, so is Y
- Geometry of Random Optimization Landscapes
  e.g.: **Kac-Rice formula**
- Statistical Query Models

Other Methods

- **Reductions** — If X is hard, so is Y
- Geometry of Random Optimization Landscapes
  e.g.: **Kac-Rice formula**
- Statistical Query Models
- Automatic algorithm learning
  — Deep Learning, **Graph Neural Networks**, ...
- . . .

Other Methods

- **Reductions** — If X is hard, so is Y
- Geometry of Random Optimization Landscapes
  e.g.: **Kac-Rice formula**
- Statistical Query Models
- Automatic algorithm learning
  — Deep Learning, **Graph Neural Networks**, ...
- . . .

Are these related?

Looking ahead

- What makes problems **hard**? How do we find $\mathbf{SNR}^*_{\mathbf{Comp}}$?

Looking ahead

- What makes problems **hard**? How do we find $\mathbf{SNR}^*_{\mathbf{Comp}}$?
- Are the **methods/heuristics** shown **equivalent/related**?

Looking ahead

- What makes problems **hard**? How do we find $\mathbf{SNR}^*_{\mathbf{Comp}}$?
- Are the **methods/heuristics** shown **equivalent/related**?
- Which atypical (random) properties are quiet? **(Hidden Rare Events)**

Looking ahead

- What makes problems **hard**? How do we find $\mathbf{SNR}^*_{\mathbf{Comp}}$?
- Are the **methods/heuristics** shown **equivalent/related**?
- Which atypical (random) properties are quiet? **(Hidden Rare Events)**
- Is there a **universality** phenomenon?

Looking ahead

- What makes problems **hard**? How do we find $\mathbf{SNR}^*_{\mathbf{Comp}}$?
- Are the **methods/heuristics** shown **equivalent/related**?
- Which atypical (random) properties are quiet? **(Hidden Rare Events)**
- Is there a **universality** phenomenon?
- Can we develop an **algorithm** that refutes all this?

Looking ahead

- What makes problems **hard**? How do we find $\mathbf{SNR}^*_{\mathbf{Comp}}$?
- Are the **methods/heuristics** shown **equivalent/related**?
- Which atypical (random) properties are quiet? **(Hidden Rare Events)**
- Is there a **universality** phenomenon?
- Can we develop an **algorithm** that refutes all this?
- Can this help **explain Learning**?

Muito Obrigado

www.afonsobandeira.com

Shameless plug: Take a look at **Ten Lectures and Forty-Two Open Problems in the Mathematics of Data Science** for some open problems