The Minimization of Random Hypergraphs

Thomas Bläsius, Tobias Friedrich, and Martin Schirneck

28th Annual European Symposium on Algorithms - September 7-9, 2020

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Universität Potsdam

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

- Unique column combinations (UCCs): entries identify the rows.

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

Add City AC

- Unique column combinations (UCCs): entries identify the rows.
- UCCs = hitting sets of hypergraph of difference sets.

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

Add City AC

Age Name Add City AC

- Unique column combinations (UCCs): entries identify the rows.
- UCCs = hitting sets of hypergraph of difference sets.

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

Add City AC

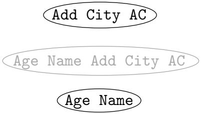
Age Name Add City AC

- Unique column combinations (UCCs): entries identify the rows.
- UCCs = hitting sets of hypergraph of difference sets.
- Non-minimal difference sets are redundant.

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

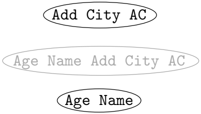


- Unique column combinations (UCCs): entries identify the rows.
- UCCs = hitting sets of hypergraph of difference sets.
- Non-minimal difference sets are redundant.

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

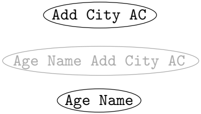


- Unique column combinations (UCCs): entries identify the rows.
- UCCs = hitting sets of hypergraph of difference sets.
- Non-minimal difference sets are redundant.

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145



- Unique column combinations (UCCs): entries identify the rows.
- UCCs = hitting sets of hypergraph of difference sets.
- Non-minimal difference sets are redundant.

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

Add City AC

Age Name Add City AC

Age Name

Age City AC

- Unique column combinations (UCCs): entries identify the rows.
- UCCs = hitting sets of hypergraph of difference sets.
- Non-minimal difference sets are redundant.

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

Add City AC

Age Name Add City AC

Age Name

Age City AC

- Unique column combinations (UCCs): entries identify the rows.
- UCCs = hitting sets of hypergraph of difference sets.
- Non-minimal difference sets are redundant.
- Discard supersets → minimization of a hypergraph.

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

Add City AC
Age Name Add City AC
Age Name
Age City AC

- Unique column combinations (UCCs): entries identify the rows.
- UCCs = hitting sets of hypergraph of difference sets.
- Non-minimal difference sets are redundant.
- Discard supersets → minimization of a hypergraph.
- Minimization is small for large databases. [Papenbrock et al. 2015] [Bläsius et al. 2019]

Data Profiling

Data profiling: mining metadata from databases.

Age	Name	Address	City	Area Code
47	Mustermann, Max	Mittelstraße 125	Potsdam	D-14467
47	Mustermann, Max	W Broadway 400	San Diego	US-CA-92101
76	Doe, John	South Street 8	London	UK-W1K
90	Nightingale, Florence	South Street 8	London	UK-W1K
25	Menigmand, Morten	Trøjburgvej 24	Aarhus	DK-8200
33	Doe, John	South Street 8	Philadelphia	US-PA-19145

Add City AC
Age Name Add City AC
Age Name
Age City AC

- Unique column combinations (UCCs): entries identify the rows.
- UCCs = hitting sets of hypergraph of difference sets.
- Non-minimal difference sets are redundant.
- Discard supersets → minimization of a hypergraph.
- Minimization is small for large databases. [Papenbrock et al. 2015] [Bläsius et al. 2019]

Why?

Random Hypergraphs

Average-case analysis needs a distribution.

Random Hypergraphs

Average-case analysis needs a distribution.

- Several random graph models exist.

Random Hypergraphs

Average-case analysis needs a distribution.

- Several random graph models exist.
- Erdős-Rényi graphs (Gilbert graphs) = "maximally random" graphs.

Random Hypergraphs

Average-case analysis needs a distribution.

- Several random graph models exist.
- Erdős-Rényi graphs (Gilbert graphs) = "maximally random" graphs.
  - $\mathcal{G}_{n,m}$ ($\mathcal{G}_{n,p}$) maximum-entropy distribution on graphs with n vertices and m edges (expected $p\binom{n}{2}$ edges).

Random Hypergraphs

Average-case analysis needs a distribution.

- Several random graph models exist.
- Erdős-Rényi graphs (Gilbert graphs) = "maximally random" graphs.
  - $\mathcal{G}_{n,m}$ ($\mathcal{G}_{n,p}$) maximum-entropy distribution on graphs with n vertices and m edges (expected $p\binom{n}{2}$ edges).
- Only few non-uniform hypergraph models. [Schmidt-Pruzan, Shamir 1985] [Chodrow 2020]

Random Hypergraphs

Average-case analysis needs a distribution.

- Several random graph models exist.
- Erdős-Rényi graphs (Gilbert graphs) = "maximally random" graphs.
  - $\mathcal{G}_{n,m}$ ($\mathcal{G}_{n,p}$) maximum-entropy distribution on graphs with n vertices and m edges (expected $p\binom{n}{2}$ edges).
- Only few non-uniform hypergraph models. [Schmidt-Pruzan, Shamir 1985] [Chodrow 2020]

Our model:

$\mathcal{B}_{n,m,p}$ maximum-entropy distribution on multi-hypergraphs with n vertices, m edges, and expected edge size pn.

Random Hypergraphs

Average-case analysis needs a distribution.

- Several random graph models exist.
- Erdős-Rényi graphs (Gilbert graphs) = “maximally random” graphs.
  - $\mathcal{G}_{n,m}$ ($\mathcal{G}_{n,p}$) maximum-entropy distribution on graphs with n vertices and m edges (expected $p\binom{n}{2}$ edges).
- Only few non-uniform hypergraph models. [Schmidt-Pruzan, Shamir 1985] [Chodrow 2020]

Our model:

$\mathcal{B}_{n,m,p}$ maximum-entropy distribution on multi-hypergraphs
with n vertices, m edges, and expected edge size pn.

Random Hypergraphs

Average-case analysis needs a distribution.

- Several random graph models exist.
- Erdős-Rényi graphs (Gilbert graphs) = "maximally random" graphs.
  - $\mathcal{G}_{n,m}$ ($\mathcal{G}_{n,p}$) maximum-entropy distribution on graphs with n vertices and m edges (expected $p\binom{n}{2}$ edges).
- Only few non-uniform hypergraph models. [Schmidt-Pruzan, Shamir 1985] [Chodrow 2020]

Our model:

$\mathcal{B}_{n,m,p}$ maximum-entropy distribution on multi-hypergraphs with n vertices, m edges, and expected edge size pn.

- Sample m subsets of $[n]$, including any vertex with probability p.

Random Hypergraphs

Average-case analysis needs a distribution.

- Several random graph models exist.
- Erdős-Rényi graphs (Gilbert graphs) = "maximally random" graphs.
  - $\mathcal{G}_{n,m}$ ($\mathcal{G}_{n,p}$) maximum-entropy distribution on graphs with n vertices and m edges (expected $p\binom{n}{2}$ edges).
- Only few non-uniform hypergraph models. [Schmidt-Pruzan, Shamir 1985] [Chodrow 2020]

Our model:

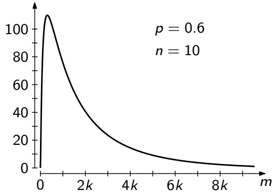
$\mathcal{B}_{n,m,p}$ maximum-entropy distribution on multi-hypergraphs
with n vertices, m edges, and expected edge size pn.

- Sample m subsets of $[n]$, including any vertex with probability p.

Expected size of the minimization of $\mathcal{B}_{n,m,p}$?

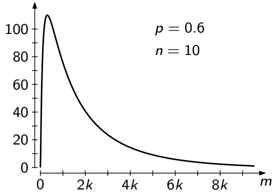
Main Results

Tight bounds on $E[|\min(\mathcal{B}_{n,m,p})|]$



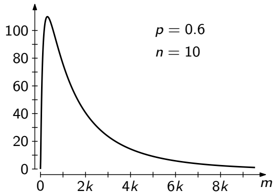
Main Results

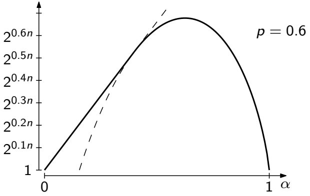
Tight bounds on $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.



Main Results

Tight bounds on $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.

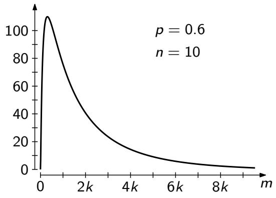


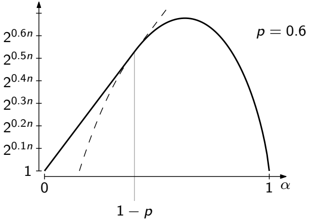


Let $\alpha = \alpha(n)$ be such that $m = 1/(1-p)^{\alpha n}$.

Main Results

Tight bounds on $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.

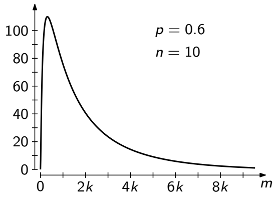


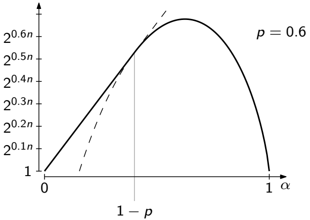


Let $\alpha = \alpha(n)$ be such that $m = 1/(1-p)^{\alpha n}$.

Main Results

Tight bounds on $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.





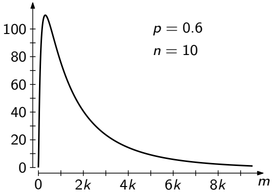
Let $\alpha = \alpha(n)$ be such that $m = 1/(1-p)^{\alpha n}$.

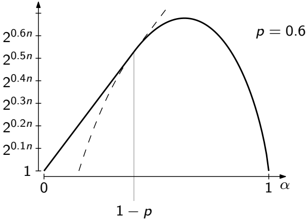
Theorem

1. If $m \leq 1/(1-p)^{(1-p)n}$, then the minimization has expected size $\Theta(m)$.

Main Results

Tight bounds on $E[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.





Let $\alpha = \alpha(n)$ be such that $m = 1/(1-p)^{\alpha n}$.

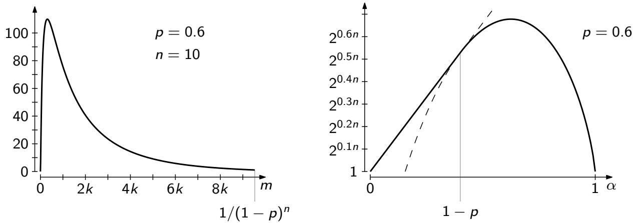
Theorem

1. If $m \leq 1/(1-p)^{(1-p)n}$, then the minimization has expected size $\Theta(m)$.
2. If $\alpha \in [1-p+\varepsilon, 1-\varepsilon]$ for any $\varepsilon > 0$, the size is $\Theta\left(2^{H(\alpha)n} \cdot p^{(1-\alpha)n} / \sqrt{n}\right)$.

($2^{H(\alpha)}$ perplexity from information theory.)

Main Results

Tight bounds on $E[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.



Let $\alpha = \alpha(n)$ be such that $m = 1/(1-p)^{\alpha n}$.

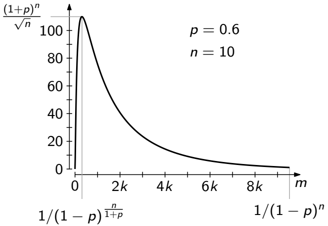
Theorem

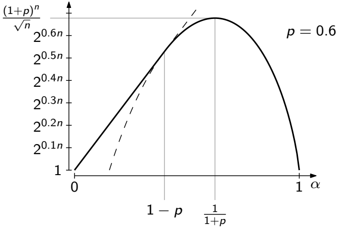
1. If $m \le 1/(1-p)^{(1-p)n}$, then the minimization has expected size $\Theta(m)$.
2. If $\alpha \in [1-p+\varepsilon, 1-\varepsilon]$ for any $\varepsilon > 0$, the size is $\Theta\left(2^{H(\alpha)n} \cdot p^{(1-\alpha)n} / \sqrt{n}\right)$.

   ($2^{H(\alpha)}$ perplexity from information theory.)
3. If $m = 1/(1-p)^{n+\omega(\log n)}$, the size is $1 + o(1)$.

Main Results

Tight bounds on $E[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.





Let $\alpha = \alpha(n)$ be such that $m = 1/(1-p)^{\alpha n}$.

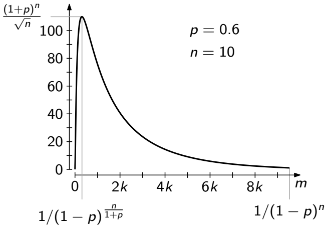
Theorem

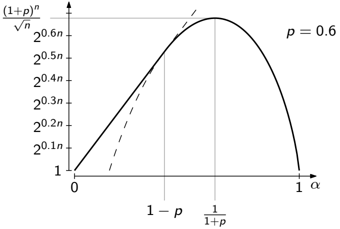
1. If $m \leq 1/(1-p)^{(1-p)n}$, then the minimization has expected size $\Theta(m)$.
2. If $\alpha \in [1-p+\varepsilon, 1-\varepsilon]$ for any $\varepsilon > 0$, the size is $\Theta\left(2^{H(\alpha)n} \cdot p^{(1-\alpha)n} / \sqrt{n}\right)$.

   ($2^{H(\alpha)}$ perplexity from information theory.)
3. If $m = 1/(1-p)^{n+\omega(\log n)}$, the size is $1 + o(1)$.

Main Results

Tight bounds on $E[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.





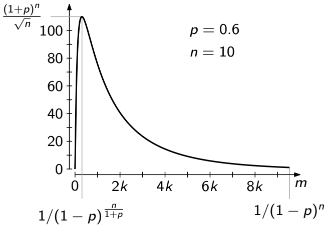
Let $\alpha = \alpha(n)$ be such that $m = 1/(1-p)^{\alpha n}$.

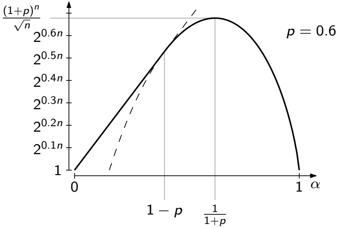
Proof recipe:

1. Establish close connection between $E[|\min(\mathcal{B}_{n,m,p})|]$ and $P[X \leq (1-\alpha)n]$ where $X \sim \text{Bin}(n,p)$.

Main Results

Tight bounds on $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.





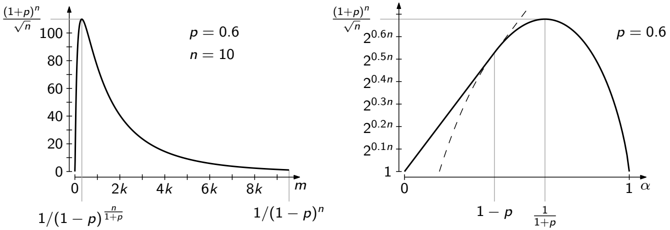
Let $\alpha = \alpha(n)$ be such that $m = 1/(1-p)^{\alpha n}$.

Proof recipe:

1. Establish close connection between $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ and $\mathrm{P}[X \le (1-\alpha)n]$ where $X \sim \mathrm{Bin}(n,p)$.
2. Improve bounds on the binomial distribution.

Main Results

Tight bounds on $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.



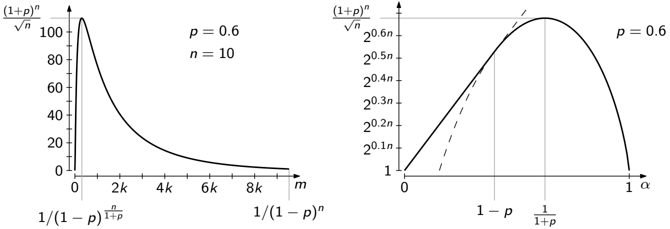
Let $\alpha = \alpha(n)$ be such that $m = 1/(1-p)^{\alpha n}$.

Proof recipe:

1. Establish close connection between $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ and $\mathrm{P}[X \leq (1-\alpha)n]$ where $X \sim \mathrm{Bin}(n, p)$.
2. Improve bounds on the binomial distribution.
3. Garnish with inequalities from combinatorics and information theory.

Main Results

Tight bounds on $E[|\min(\mathcal{B}_{n,m,p})|]$ & phase transition at $m^* = 1/(1-p)^{(1-p)n}$.



Let $\alpha = \alpha(n)$ be such that $m = 1/(1-p)^{\alpha n}$.

Proof recipe:

1. Establish close connection between $E[|\min(\mathcal{B}_{n,m,p})|]$ and $P[X \leq (1-\alpha)n]$ where $X \sim \text{Bin}(n,p)$.
2. Improve bounds on the binomial distribution.
3. Garnish with inequalities from combinatorics and information theory.

The Chernoff–Hoeffding Theorem

How likely $X \sim \mathrm{Bin}(n, p)$ deviates from $\mathrm{E}[X] = pn$?

The Chernoff–Hoeffding Theorem

How likely $X \sim \mathrm{Bin}(n, p)$ deviates from $\mathrm{E}[X] = pn$?

Estimate $\mathrm{P}[X \leq xn]$ for $x \leq p$. (Via Kullback–Leibler divergence $\mathrm{D}(x \,\|\, p)$.)

The Chernoff–Hoeffding Theorem

How likely $X \sim \mathrm{Bin}(n, p)$ deviates from $\mathrm{E}[X] = pn$?

Estimate $\mathrm{P}[X \leq xn]$ for $x \leq p$. (Via Kullback–Leibler divergence $\mathrm{D}(x \,\|\, p)$.)

- Chernoff–Hoeffding theorem: $\mathrm{P}[X \leq xn] \leq 2^{-\mathrm{D}(x \,\|\, p)n}$. [Hoeffding 1963]

The Chernoff–Hoeffding Theorem

How likely $X \sim \mathrm{Bin}(n, p)$ deviates from $\mathrm{E}[X] = pn$?

Estimate $\mathrm{P}[X \leq xn]$ for $x \leq p$. (Via Kullback–Leibler divergence $\mathrm{D}(x \,\|\, p)$.)

- Chernoff–Hoeffding theorem: $\mathrm{P}[X \leq xn] \leq 2^{-\mathrm{D}(x \,\|\, p)n}$. [Hoeffding 1963]
- Stirling's approximation: $\mathrm{P}[X \leq xn] \geq \frac{1}{\sqrt{8nx(1-x)}} \cdot 2^{-\mathrm{D}(x \,\|\, p)n}$.

The Chernoff–Hoeffding Theorem

How likely $X \sim \mathrm{Bin}(n,p)$ deviates from $\mathrm{E}[X] = pn$?

Estimate $\mathrm{P}[X \leq xn]$ for $x \leq p$. (Via Kullback–Leibler divergence $\mathrm{D}(x \,\|\, p)$.)

- Chernoff–Hoeffding theorem: $\mathrm{P}[X \leq xn] \leq 2^{-\mathrm{D}(x \,\|\, p)n}$. [Hoeffding 1963]
- Stirling's approximation: $\mathrm{P}[X \leq xn] \geq \frac{1}{\sqrt{8nx(1-x)}} \cdot 2^{-\mathrm{D}(x \,\|\, p)n}$.
- Closing the $\mathrm{O}(\sqrt{n})$-gap down to constants.

The Chernoff–Hoeffding Theorem

How likely $X \sim \mathrm{Bin}(n, p)$ deviates from $\mathrm{E}[X] = pn$?

Estimate $\mathrm{P}[X \leq xn]$ for $x \leq p$. (Via Kullback–Leibler divergence $\mathrm{D}(x \,\|\, p)$.)

- Chernoff–Hoeffding theorem: $\mathrm{P}[X \leq xn] \leq 2^{-\mathrm{D}(x \,\|\, p) n}$. [Hoeffding 1963]
- Stirling's approximation: $\mathrm{P}[X \leq xn] \geq \frac{1}{\sqrt{8nx(1-x)}} \cdot 2^{-\mathrm{D}(x \,\|\, p) n}$.
- Closing the $\mathrm{O}(\sqrt{n})$-gap down to constants.

Theorem: Tight Chernoff-Hoeffding.

If $0 < x < p$ is a constant, then $\mathrm{P}[X \leq xn] = \Theta\big(2^{-\mathrm{D}(x \,\|\, p) n} / \sqrt{n}\big)$.

The same holds if $x = x(n) \in [\varepsilon, p - \varepsilon]$ for any $\varepsilon > 0$.

The Chernoff–Hoeffding Theorem

How likely $X \sim \mathrm{Bin}(n, p)$ deviates from $\mathrm{E}[X] = pn$?

Estimate $\mathrm{P}[X \leq xn]$ for $x \leq p$. (Via Kullback–Leibler divergence $\mathrm{D}(x \,\|\, p)$.)

- Chernoff–Hoeffding theorem: $\mathrm{P}[X \leq xn] \leq 2^{-\mathrm{D}(x \,\|\, p)n}$. [Hoeffding 1963]
- Stirling's approximation: $\mathrm{P}[X \leq xn] \geq \frac{1}{\sqrt{8nx(1-x)}} \cdot 2^{-\mathrm{D}(x \,\|\, p)n}$.
- Closing the $O(\sqrt{n})$-gap down to constants.

Theorem: Tight Chernoff-Hoeffding.

If $0 < x < p$ is a constant, then $\mathrm{P}[X \leq xn] = \Theta\left(2^{-\mathrm{D}(x \,\|\, p)n} / \sqrt{n} \right)$.

The same holds if $x = x(n) \in [\varepsilon, p - \varepsilon]$ for any $\varepsilon > 0$.

Combine bound on ratio $\mathrm{P}[X \leq xn] / \mathrm{P}[X = xn]$ [Klar 2000]

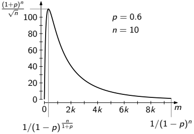
with perplexity $\binom{n}{xn} = \Theta(2^{\mathrm{H}(x)n} / \sqrt{n})$.

Summary

- Analysis of minimization of maximum-entropy multi-hypergraph $\mathcal{B}_{n,m,p}$.

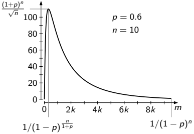
Summary

- Analysis of minimization of maximum-entropy multi-hypergraph $\mathcal{B}_{n,m,p}$.
- Implications for large databases.

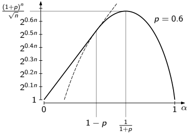


Summary

- Analysis of minimization of maximum-entropy multi-hypergraph $\mathcal{B}_{n,m,p}$.
- Implications for large databases.

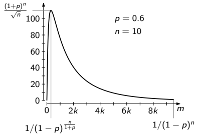


- Phase transition at $m^* = 1/(1-p)^{(1-p)n}$.

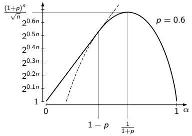


Summary

- Analysis of minimization of maximum-entropy multi-hypergraph $\mathcal{B}_{n,m,p}$.
- Implications for large databases.



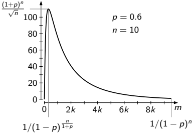
- Phase transition at $m^* = 1/(1-p)^{(1-p)n}$.



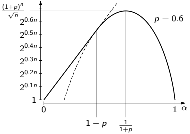
- Tight bounds on binomial distribution $\mathrm{P}[X \le xn] = \Theta\left(2^{-\mathrm{D}(x \,\|\, p)n} / \sqrt{n}\right)$.

Summary

- Analysis of minimization of maximum-entropy multi-hypergraph $\mathcal{B}_{n,m,p}$.
- Implications for large databases.



- Phase transition at $m^* = 1/(1-p)^{(1-p)n}$.



- Tight bounds on binomial distribution $\mathrm{P}[X \le xn] = \Theta\left(2^{-\mathrm{D}(x \,\|\, p)n} / \sqrt{n}\right)$.

Thank you.