AN ASYMPTOTIC THEORY FOR K-DOMINANT SKYLINES OF RANDOM SAMPLES

Hsien-Kuei Hwang, ***Academia Sinica, Taiwan***
(joint with Wei-Mei Chen, Tsugn-Hsi Tsai)
March 11, 2011

Le Groupe de Travail AL É A

SKYLINE

According to Merriam-Webster

An outline (as of buildings or a mountain range) against the background of the sky

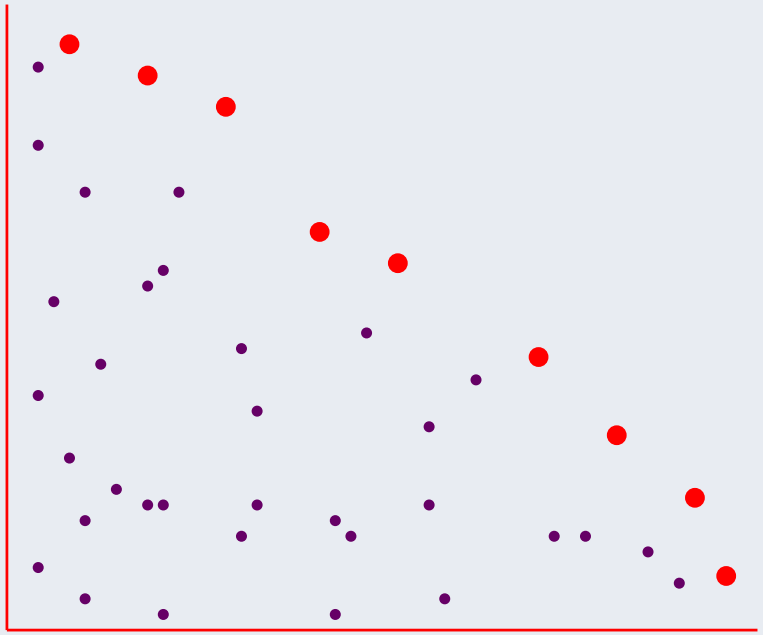
DOMINANCE

$$\mathbf{x} = (x_1, \ldots, x_d) \textbf{ \textit{dominates} } \mathbf{y} = (y_1, \ldots, y_d)$$

$$\textbf{if } x_i > y_i \textbf{ for } i = 1, \ldots, d.$$

MAXIMA

Nondominated points in a point set are called *maximal points* or *maxima* of the point set.



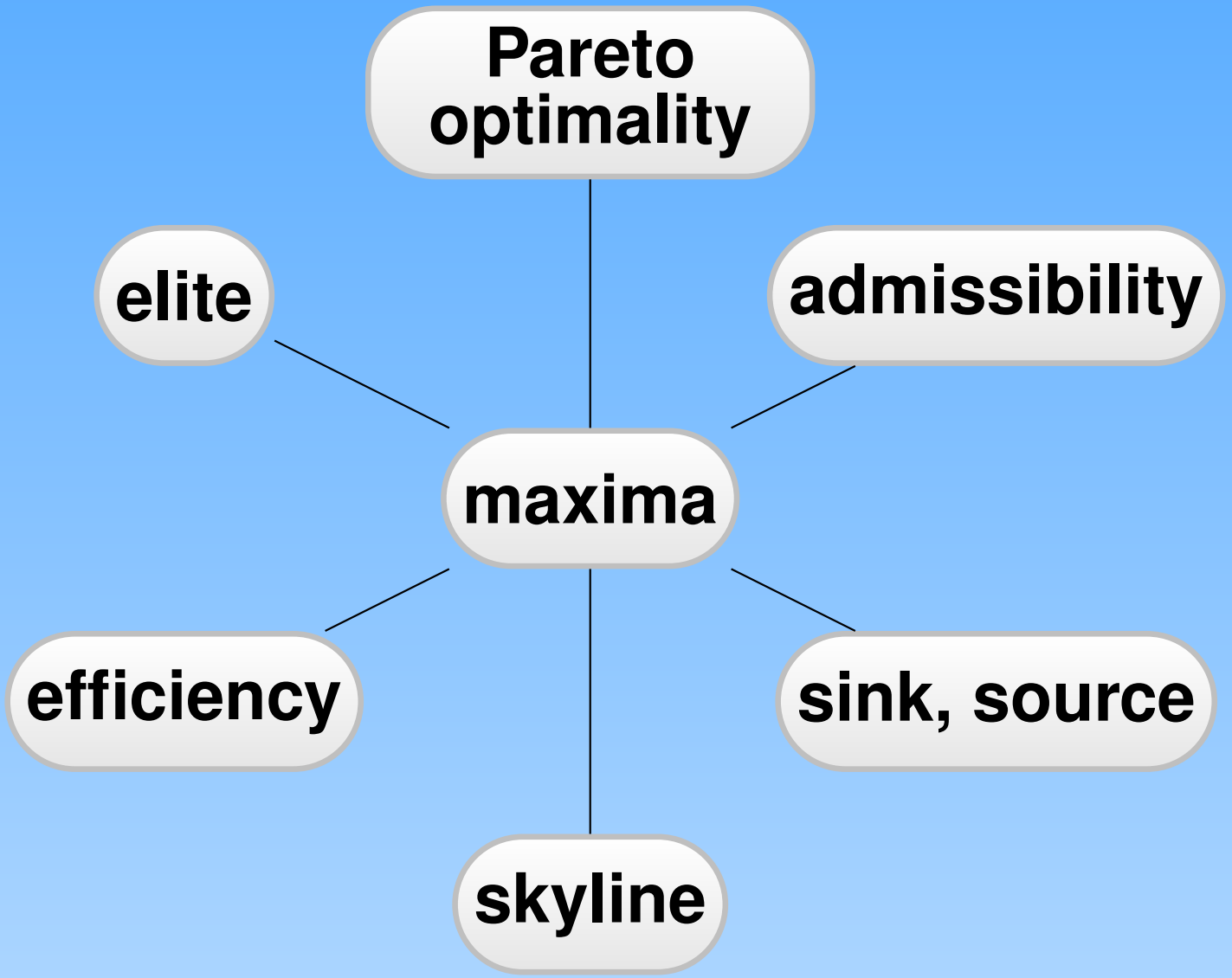
SKYLINE = MINIMA

$\mathbf{x} = (x_1, \ldots, x_d)$ ***dominates*** $\mathbf{y} = (y_1, \ldots, y_d)$

if $x_i < y_i$ **for** $i = 1, \ldots, d$.

Nondominated points in a point set are called *minimal points* or *minima* of the point set.

SYNONYMS OF MAXIMA



SYNONYMS AND UNDERLYING AREAS

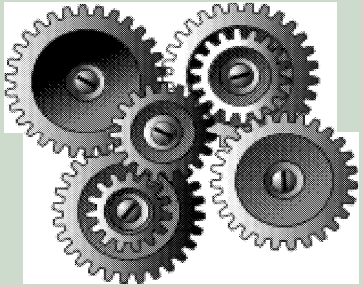
Diverse terms show its popularity

- ***Pareto optimality*** **(Pareto front, Pareto efficiency, Pareto solutions, . . .) in multiobjective optimization or multicriteria decision**
- ***Efficiency*** **in econometrics**
- ***Sink or source*** **in applied probability**
- ***Admissibility*** **in decision theory**
- ***Elite*** **in evolutionary computation**
- ***Skyline*** **in databases = minima**

APPLICATIONS

USEFULNESS

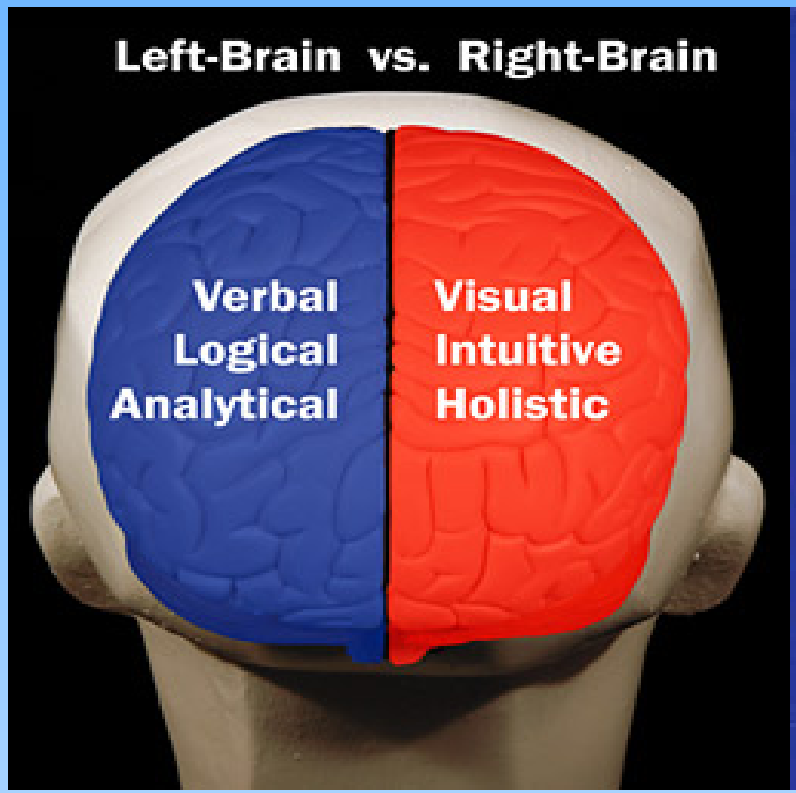
- **Political science**
- **Social science**
- **Engineering**
- **Econometrics**
- **Operations Research**
- **Biology**
- **Education**
- **Computer algorithms**

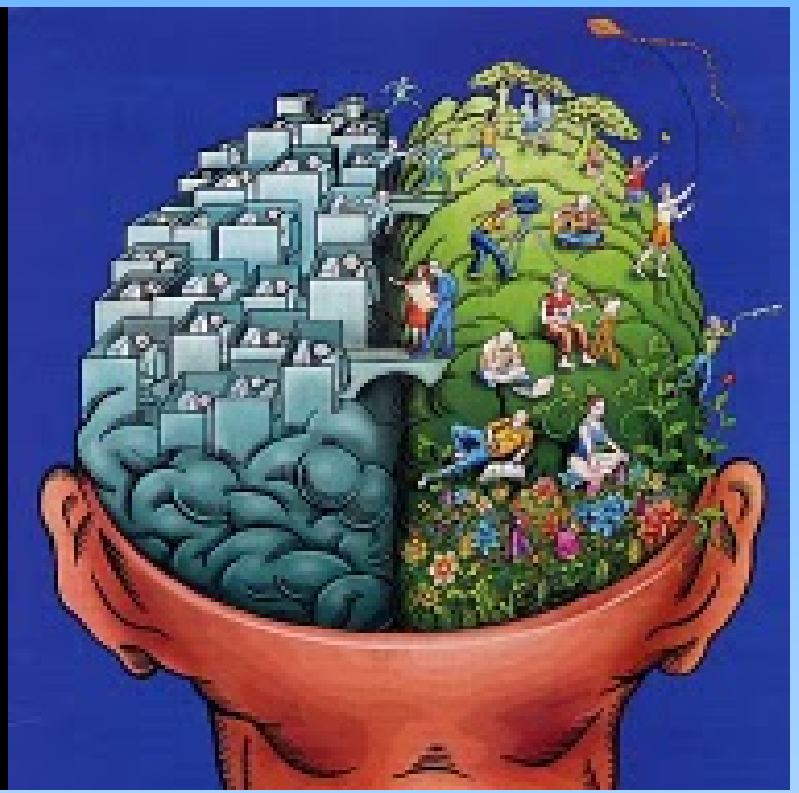


TRADEOFFS EVERYWHERE

Daily life is full of tradeoffs

To eat or not to eat, to buy or to sell (stock shares)
Eat this or eat that, buy this or buy that, . . .





How to balance is an optimization problem

Efficiency is equivalent to Pareto optimality

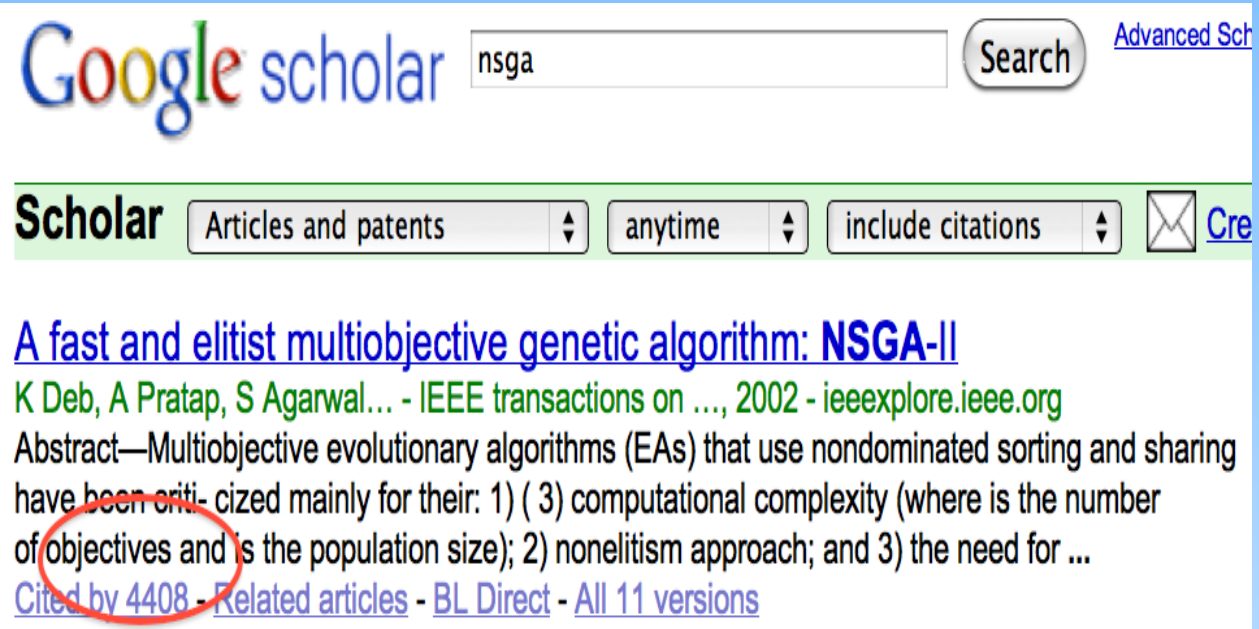
Important issue: description and characterization of Pareto optimal solutions.

Vilfredo Pareto (1848-1923)

ELITISM IN MULTIOBJECTIVE EVOLUTIONARY ALGORITHM

Evolutionary algorithm (from Wiki)

A subset of evolutionary computation, a generic population-based metaheuristic optimization algorithm. An EA uses some mechanisms inspired by biological evolution: reproduction, mutation, recombination, and selection.



EREWHON'S KING PROBLEM

Preparata and Shamos (1985): *Comput. Geo.*, p. 151

4.1.3 The problem of the maxima of a point set

This section is devoted to the discussion of a problem that bears an intriguing resemblance to that of the convex hull and yet on a deeper level, to be elucidated later, is fundamentally different from it. This problem occurs in a large number of applications, in statistics, economics, operations research, etc. Indeed, it was originally described as the "floating-currency problem"[5]:

In Erehwon, every citizen has a portfolio of foreign currencies; these foreign currencies wildly fluctuate in their values, so that every evening a person whose portfolio has the largest cumulative value is declared the King of Erehwon. Which is the smallest subset of the population that is certain to contain all potential kings?

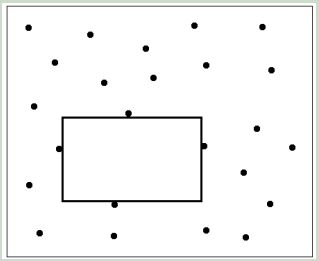
Finding the maxima

- a prototype problem with many algorithmic and practical applications;
- a component problem in several applications;
- related to many computational geometric problems.

A fundamental problem.

Examples

- **the minimum independent dominating set in permutation graphs**
- **the largest empty rectangle**



- **restricted empty rectangles**
- **the enclosure problem for planar (rectilinear) d-gon (which in turn has several applications in CAD systems for VLSI circuits).**
- **Other examples include: convex hull, polygon decomposition, shortest path problem, finding empty slimplices, geometric containment problem, official statistics, grid placement problem, . . .**

SKYLINE IN DATABASES

Introduced by Börzsönyi, Kossmann and Stocker (2001)

Travelers looking for cheaper hotels closer to the beach.

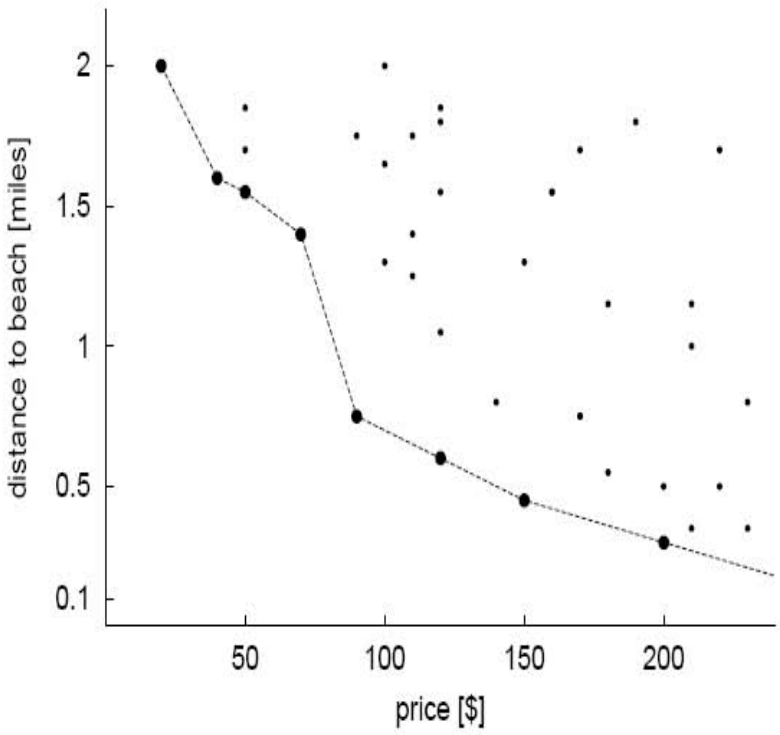


Figure 1: Skyline of Hotels

Figure 2: Skyline of Manhattan

SKYLINE OPERATOR

Google scholar | skyline operator | Search | Advanced Scholar Search

Scholar | Articles excluding patents | anytime | include citations | Create email

The skyline operator
S Borzsony, D Kossmann... - Data Engineering, 2001. ..., 2002 - ieeexplore.ieee.org
We propose to extend database systems by a **Skyline** op- eration. This operation $filters out a set of interesting points from a potentially large set of data points. A point is inter- esting if it is not dominated by any other point. For example, a hotel might be interesting for somebody ...
Cited by 456 - Related articles - Check for it @ INRIA - BL Direct - All 9 versions - Import into BibTeX

[PDF] **The skyline operator**
S Börzsönyi, D Kossmann... - Proceedings of the 17th ..., 2001 - Citeseer
We propose to extend database systems by a **Skyline** operation. This operation filters out a set of interesting points from a potentially large set of data points. A point is interesting if it is not dominated by any other point. For example, a hotel might be interesting for somebody traveling to ...
Cited by 444 - Related articles - View as HTML - Check for it @ INRIA - All 17 versions - Import into BibTeX

MINIMAL DIRECTED SPANNING TREES ON RANDOM POINTS

Bhatt and Roy (2004), Bai, Lee and Penrose (2006)

Motivation: information transmission through wireless networks, hydrology models

On a random directed spanning tree SGSA • 21

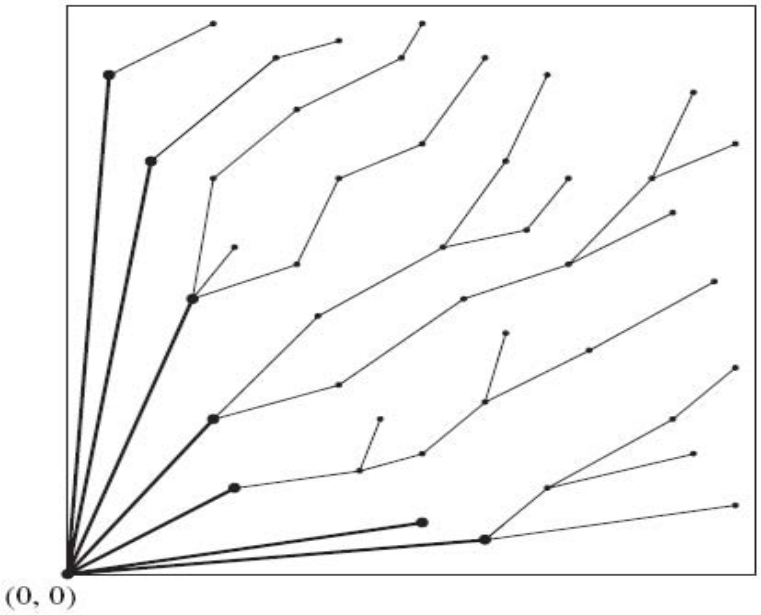
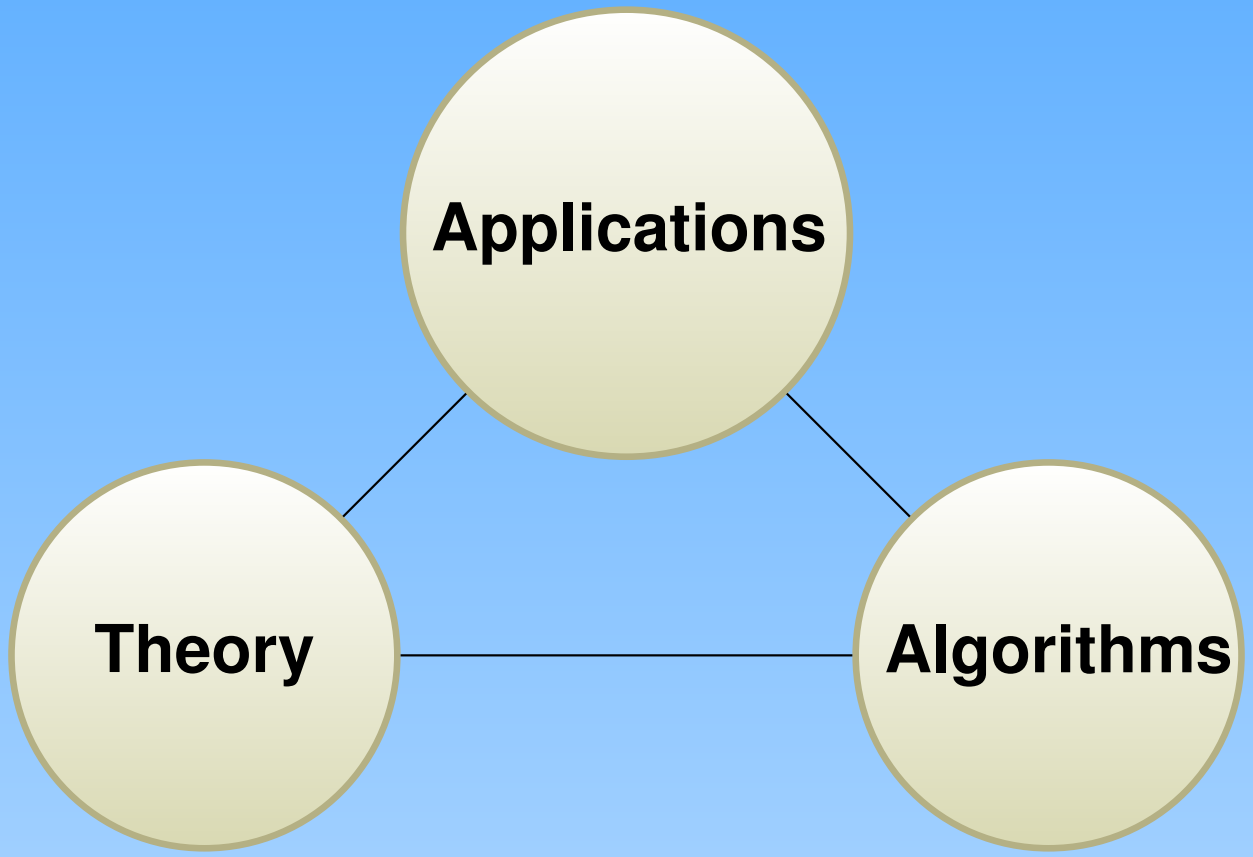
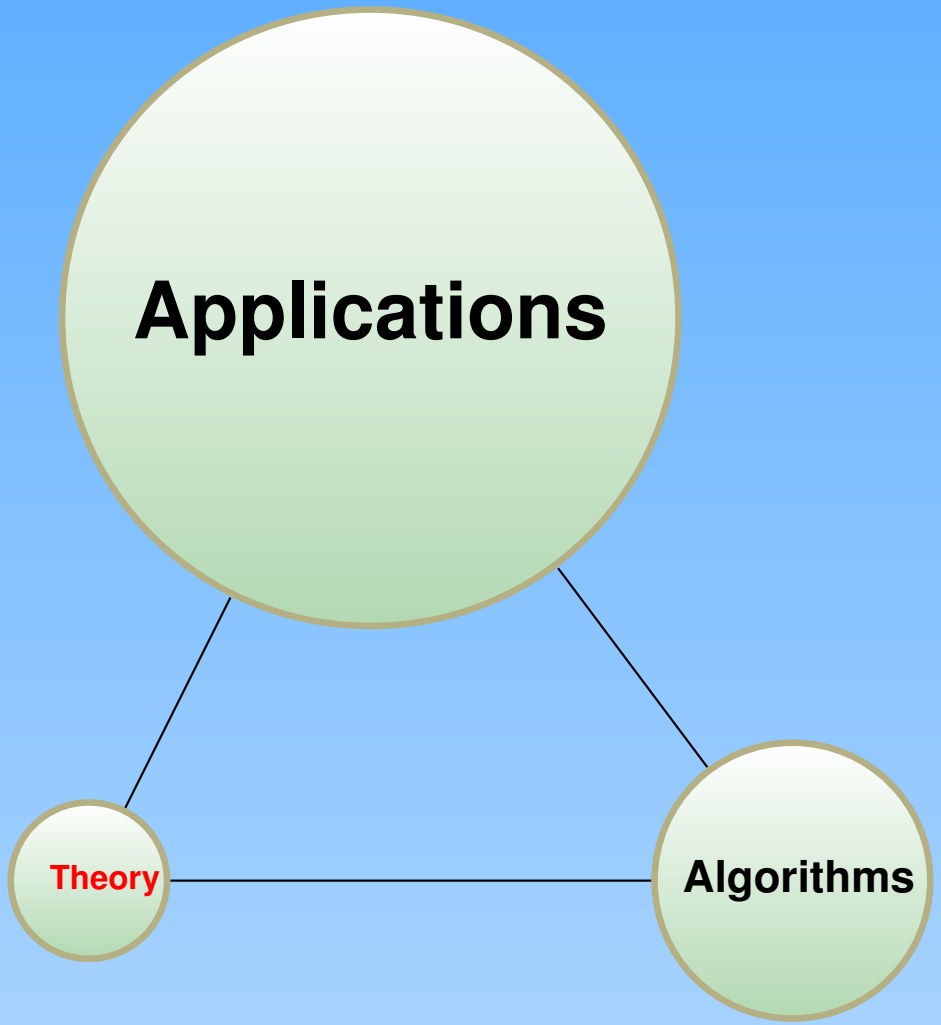


FIGURE 1: A minimal directed spanning tree. The subgraph formed by the bold edges is L_n.

MULTIFACETED ASPECTS OF MAXIMA



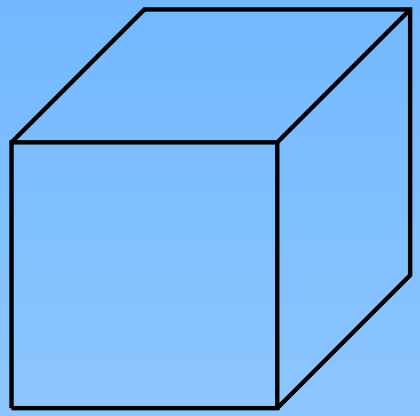
MULTIFACETED ASPECTS OF MAXIMA

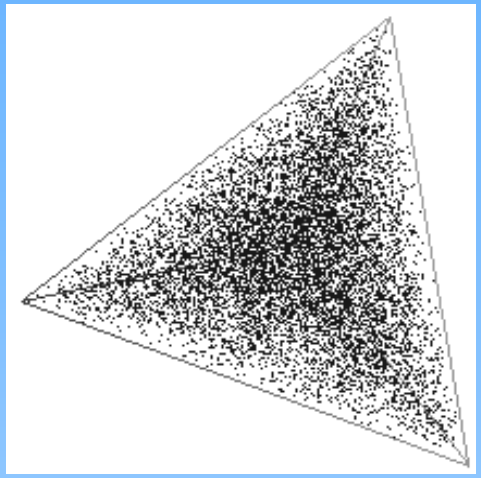


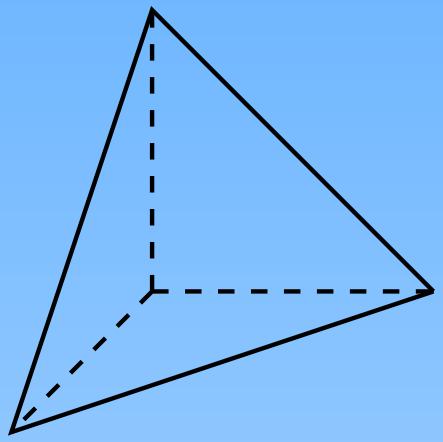
PROBABILISTIC THEORY OF MAXIMA

n points randomly sampled from $\mathscr{D}$

How many maxima?







Commonly used models: hypercube and d-simplex

n points uniformly and independently from $\mathscr{D} = [0,1]^d$ or from the d-dimensional simplex

$$\{x_1 + \cdots + x_d \leqslant 1, x_1, \ldots, x_d \in [0,1]\}$$

KNOWN RESULTS FOR MAXIMA IN HYPERCUBES

$M_{n,d} := \#(\text{maxima of } \{\mathbf{p}_1, \ldots, \mathbf{p}_n\})$

Let $\mu_{n,d} := \mathbb{E}(M_{n,d})$. Barndorff-Nielsen & Sobel (1966)
For $d \geqslant 2$,

$$\mu_{n,d} = \sum_{1 \leqslant k \leqslant n} \binom{n}{k} (-1)^{k-1} k^{1-d}$$
$$= \frac{(\log n)^{d-1}}{(d-1)!} \left(1 + O\left((\log n)^{-1}\right)\right).$$

Rederived or simplified in

Berezovskii and Travkin (1975), Ivanin (1975), Bentley et al. (1978), Devroye (1980), O'Neill (1981), Buchta (1989)

SUMMATION FORMULAE FOR $\mu_{n,d}$

$$H_n^{(a)} := \sum_{1\leqslant j\leqslant n} 1/j^a$$

$$\mu_{n,d} = \sum_{1\leqslant k\leqslant n} \binom{n}{k} (-1)^{k-1} k^{1-d},$$

$$\mu_{n,d} = \sum_{1\leqslant i_1\leqslant\cdots\leqslant i_{d-1}} \frac{1}{i_1\cdots i_{d-1}},$$

$$\mu_{n,d} = \sum_{\substack{i_1+2i_2+\cdots(d-1)i_{d-1}=d-1\\ i_1,\ldots,i_{d-1}\geqslant 0}} \frac{H_n^{i_1}(H_n^{(2)})^{i_2}\cdots(H_n^{(d-1)})^{i_{d-1}}}{i_1!\cdots i_{d-1}!1^{i_1}\cdots(d-1)^{i_{d-1}}}.$$

INTEGRAL REPRESENTATIONS FOR $\mu_{n,d}$

$$\mu_{n,d} = n \int_{(0,1)^d} (1 - x_1 x_2 \cdots x_d)^{n-1} \, \mathrm{d}\mathbf{x},$$

$$\mu_{n,d} = \frac{n}{(d-1)!} \int_0^1 (1-x)^{n-1} (-\log x)^{d-1} \, \mathrm{d}x,$$

$$\mu_{n,d} = \frac{1}{2\pi i} \oint_{|z|=r<1} z^{-d} \prod_{1 \leqslant j \leqslant n} \frac{1}{1 - z/j} \, \mathrm{d}z,$$

$$\mu_{n,d} = \frac{(-1)^n}{2\pi i} \int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} \frac{n!}{s^d (s-1) \cdots (s-n)} \, \mathrm{d}s.$$

RECURRENCE RELATIONS OF $\mu_{n,d}$

For $n \geqslant 1$ and $d \geqslant 2$

$$\mu_{n,d} = \mu_{n-1,d} + \frac{\mu_{n,d-1}}{n},$$

$$\mu_{n,d} = \sum_{1 \leqslant j \leqslant n} \frac{\mu_{j,d-1}}{j},$$

$$\mu_{n,d} = \frac{1}{d-1} \sum_{1 \leqslant j \leqslant d-1} H_n^{(d-j)} \mu_{n,j},$$

PROBABILITY EXPRESSIONS FOR $\mu_{n,d}$

$$\mu_{n,d} = n\mathbb{E}\left[(1 - U_1 U_2 \cdots U_d)^{n-1}\right],$$
$$\mu_{n,d} = n\mathbb{P}(G_2 + \cdots + G_n < d),$$

where $U_1, U_2, \ldots, U_d$ **are uniform**$[0, 1]$ **RVs and**

$$\mathbb{E}[z^{G_j}] = \frac{1 - 1/j}{1 - z/j} \qquad (2 \leqslant j \leqslant n);$$

Flajolet et al. (1995): $\frac{\mu_{n,d}}{n}$ **is the probability that the first subtree in a random quadtree of** n **nodes is empty.**

VARIANCE OF $M_{n,d}$

Bai, Chao, H., Liang (1998)

For $d \geqslant 2$,

$$\frac{\mathbb{V}(M_{n,d})}{(\log n)^{d-1}} = \frac{1}{(d-1)!} + \kappa_d + O\left((\log n)^{-1}\right).$$

where

$$\kappa_d = \frac{1}{(d-1)!} \sum_{\ell \geqslant 1} \frac{1}{\ell^2} \sum_{1 \leqslant p,q \leqslant \ell} \binom{\ell}{p} \binom{\ell}{q} (-1)^{p+q} pq$$
$$\times \left(\left(p^{-1} + q^{-1}\right)^{d-1} - p^{1-d} - q^{1-d} \right).$$

Bai, Devroye, H., Tsai (2005): an asymptotic expansion.

κ_d FOR $2 \leqslant d \leqslant 9$

$\zeta(s)$: Riemann's zeta function

$$
\begin{aligned}
\kappa_2 &= 0, \\
\kappa_3 &= \zeta(2), \\
\kappa_4 &= 2\zeta(3), \\
\kappa_5 &= \tfrac{33}{16}\,\zeta(4), \\
\kappa_6 &= \tfrac{5}{4}\zeta(5) + \tfrac{1}{6}\zeta(2)\zeta(3), \\
\kappa_7 &= \tfrac{1451}{1728}\,\zeta(6) + \tfrac{7}{72}\zeta(3)^2, \\
\kappa_8 &= \tfrac{1729}{5760}\,\zeta(7) + \tfrac{13}{360}\,\zeta(2)\zeta(5) + \tfrac{181}{1440}\,\zeta(3)\zeta(4), \\
\kappa_9 &= \tfrac{1891}{89600}\,\zeta(2)^4 + \tfrac{\zeta(2)\zeta(3)^2}{320} + \tfrac{11}{160}\,\zeta(3)\zeta(5) \\
&\quad - \tfrac{17}{1920}\sum_{k\geqslant 1} k^{-2} \sum_{j>k} j^{-6}.
\end{aligned}
$$

Bai et al. (1998, 2005): $\{\kappa_d\}_{2\leqslant d\leqslant 8}$;
Costermans's PhD Thesis (2008): $\{\kappa_d\}_{9\leqslant d\leqslant 13}$.

REPRESENTATIONS FOR κ_d

$$\begin{aligned}
\kappa_d &= \frac{1}{(d-1)!}\sum_{j\geqslant 1}\frac{1}{j^2}\sum_{1\leqslant m,k\leqslant j}\binom{j}{m}\binom{j}{k}(-1)^{m+k}mk\left(\left(m^{-1}+k^{-1}\right)^{d-1}-m^{1-d}-k^{1-d}\right)\\
&= \sum_{j\geqslant 1}j^{-2}\sum_{1\leqslant m\leqslant d-2}\frac{\mu_{j,m}\mu_{j,d-1-m}}{m!(d-1-m)!}\\
&= -\frac{2}{(d-1)!^2}\int_0^1\frac{(-\log x)^{d-1}}{(1+x)^2}\,\mathrm{d}x+\frac{1}{(d-1)!}\sum_{1\leqslant k<d}\binom{d}{k}\frac{1}{(k-1)!(d-1-k)!}\\
&\quad\times\int_0^1\!\!\int_0^1(-\log x)^{k-1}(-\log y)^{d-1-k}\left(\frac{1}{(x+y-xy)^2}-\frac{1}{(x+y)^2}\right)\mathrm{d}x\,\mathrm{d}y\\
&= \sum_{1\leqslant k\leqslant d-2}\frac{1}{k!(d-1-k)!(k-1)!(d-2-k)!}\int_0^1\!\!\int_0^1\frac{(-\log x)^{k-1}(-\log y)^{d-2-k}}{x+y-xy}\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}$$

For others, see Costermans (2008), Costermans and Minh (2008).

CENTRAL LIMIT THEOREM FOR $M_{n,d}$

Baryshnikov (2000)

A CLT for $M_{n,d}$.

Bai, Devroye, H., Tsai (2005)

Prove a CLT for $M_{n,d}$ by *Poisson processes* and *Stein's method* with a rate

$$O\left((\log n)^{-(d-1)/4}(\log\log n)^{d}\right).$$

Open

Prove the Berry-Esseen bound $O((\log n)^{-(d-1)/2})$

Same theory can be developed for multivariate records

THE CURSE OF DIMENSIONALITY

Phase transition at $d = \log n$

$$\mu_{n,d} \sim \begin{cases} \Gamma\left(1 - \dfrac{d-1}{\log n}\right) \dfrac{(\log n)^{d-1}}{(d-1)!}, & \textbf{if } \dfrac{d - \log n}{\sqrt{\log n}} \to -\infty, \\ \Phi\left(\dfrac{d - \log n}{\sqrt{\log n}}\right) n, & \textbf{if } \dfrac{d - \log n}{\sqrt{\log n}} = O(1), \\ n, & \textbf{if } \dfrac{d - \log n}{\sqrt{\log n}} \to \infty, \end{cases}$$

where $\Gamma =$ the Gamma function and $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-t^2/2} \mathrm{d}t$ the standard normal distribution function.

TOO MANY MAXIMA TO USE

Approximate values of $\mu_{n,d}$

n \ d	4	5	6	7
10^4	164.7	426.3	902.7	1633.1
10^5	304.9	955.8	2432.1	5239.4

Other alternatives are needed

k-DOMINANT SKYLINES

Skylines = Minima

$$\mathbf{x} = (x_1, \ldots, x_d) \textbf{ dominates } \mathbf{y} = (y_1, \ldots, y_d)$$

$$\textbf{if } x_i < y_i \textbf{ for } i = 1, \ldots, d.$$

Idea: change full dominance to partial

The point x *k-dominates* y if there are k ($1 \leqslant k \leqslant d$) dimensions in which x_j is not greater than y_j and is less than in at least one of these k dimensions.

The points that are not k-dominated by any other points define the *k-dominant skyline*.

k-DOMINANT SKYLINES

proposed by Chan-Jagadish-Tan-Tung-Zhang, 2006

Google scholar "k-dominant skylines"

Scholar Articles and patents anytime include cita

Finding **k-dominant skylines** in high dimensional space

CY Chan, HV Jagadish, KL Tan… - Proceedings of the …, 2006 - portal.acm

ABSTRACT Given a d-dimensional data set, a point p dominates another poi
than or equal to q in all dimensions and better than q in at least one dimensio
point if there does not exists any point that can dominate it. Skyline queries, w

Cited by 143 - Related articles - Check for it @ INRIA - All 25 versions

k-DOMINANT SKYLINES

Google scholar

Search

Advanced Scholar S

Search within articles citing Chan: Finding k-dominant skyline

Scholar All anytime include citations Create em

Probabilistic skylines on uncertain data
J Pei, B Jiang, X Lin… - … conference on very large data bases, 2007 - portal.acm.org
ABSTRACT Uncertain data are inherent in some important applications. Although a considerable amount of research has been dedi- cated to modeling uncertain data and answering some types of queries on uncertain data, how to conduct advanced anal- ysis on uncertain data …
Cited by 150 - Related articles - All 14 versions

Selecting stars: The k most representative skyline operator
X Lin, Y Yuan, Q Zhang… - Data Engineering, 2007. …, 2007 - ieeexplore.ieee.org
Page 1. Selecting Stars: The k Most Representative Skyline Operator Xuemin Lin 1 Yidong Yuan 1 Qing Zhang 2 Ying Zhang 1 1 School of Computer Science and Engineering 2 E-Health Research Center The University of New South Wale & NICTA, Australia …
Cited by 106 - Related articles - Check for it @ INRIA - All 8 versions

Efficient skyline computation over low-cardinality domains
M Morse, JM Patel… - Proceedings of the 33rd …, 2007 - portal.acm.org
ABSTRACT Current skyline evaluation techniques follow a common paradigm that eliminates data elements from skyline consideration by find- ing other elements in the dataset that dominate them. The perfor- mance of such techniques is heavily influenced by the underlying data …
Cited by 82 - Related articles - All 9 versions

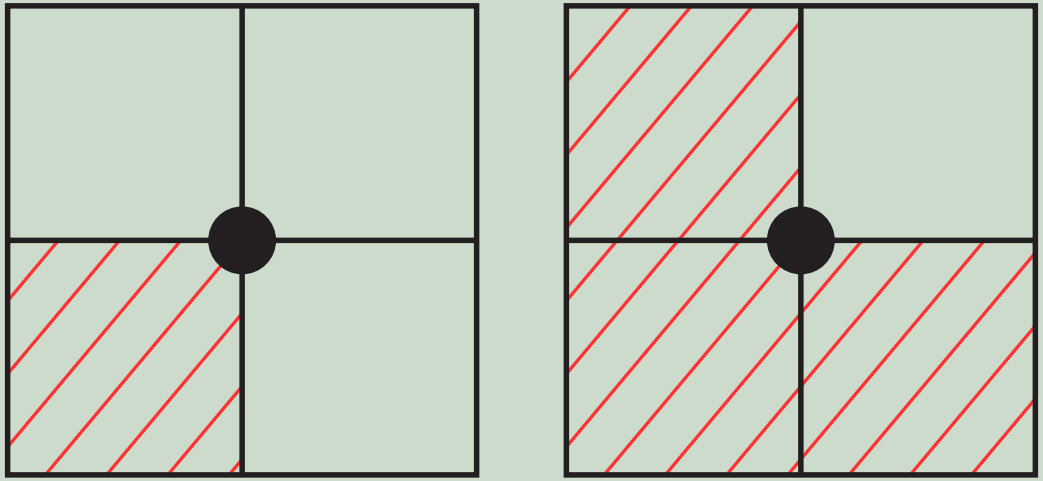
Approaching the skyline in Z order
KCK Lee, B Zheng, H Li… - Proceedings of the 33rd …, 2007 - portal.acm.org
ABSTRACT Given a set of multidimensional data points, skyline query retrieves a set of data points that are not dominated by any other points. This query is useful for multi-preference analysis and decision making. By analyzing the skyline query, we observe a close connection be- …
Cited by 79 - Related articles - All 7 versions

MOTIVATION

Monotonicity of $M_{d,k}(n)$ in k

$$M_{d,1}(n) \leqslant M_{d,2}(n) \leqslant \cdots \leqslant M_{d,d-1}(n) \leqslant M_{d,d}(n)$$

$d = 2$: $k = 2$ (left) and $k = 1$ (right)



AN ASYMPTOTIC VANISHING PROPERTY

Theorem. ($[0,1]^d$, uniform, independent)

Let $M_{d,k}(n) := \#$ ***points in the*** k***-dominant skyline of a random sample. Then***

$$\mathbb{E}[M_{d,k}(n)] \to 0 \qquad (1 \leqslant k < d; d \geqslant 2),$$

as $n \to \infty$ ***and*** $d = O(1)$***.***

AN ASYMPTOTIC VANISHING PROPERTY

Idea of proof

$$\begin{aligned}
&\mathbb{E}[M_{d,d-1}(n)] \\
&\quad = n\mathbb{P}(\mathbf{p}_1 \textbf{ is a } (d-1)\textbf{-dominant skyline point}) \\
&\quad = n\int_{[0,1]^d} (1 - |B_{d-1}(\mathbf{x})|)^{n-1}\, \mathbf{dx},
\end{aligned}$$

where $B_k(\mathbf{x})$ denotes the region of the points in $[0,1]^d$ that k-dominate $\mathbf{x}$.

THE ZERO-INFINITY PROPERTY

$$\mathbb{E}[M_{d,k}(n)] \to \begin{cases} 0, & \textbf{if } 1 \leqslant k < d; \\ \infty, & \textbf{if } k = d. \end{cases}$$

THE ZERO-INFINITY PROPERTY

$$\mathbb{E}[M_{d,k}(n)] = \begin{cases} O\left(n^{1-d/k}\right), & \textbf{if } 1 \leqslant k < d; \\ O\left(\dfrac{(\log n)^{d-1}}{(d-1)!}\right), & \textbf{if } k = d. \end{cases}$$

Lose too many from full to partial dominance

ALLOWING MORE POINTS

Theorem. ($[0,1]^d$, uniform, independent)

Let $L_{d,k}(n,j) := \#$ points in the random sample that are k-dominated by exactly j points ($L_{d,k}(n,0) = M_{d,k}(n)$).

$$\mathbb{E}[L_{d,k}(n,j)] \to 0 \qquad (1 \leqslant k < d; d = O(1)),$$

uniformly for $j = O(n^{(1-\varepsilon)/d})$, as $n \to \infty$.

Useless even allowing more points

same zero-infinity property holds

CHANGING MODEL: HYPERCUBE $\to$ d-SIMPLEX

Theorem. (d-simplex, uniform, independent)

Let $M_{d,k}^{[s]}(n) := \#$ points in the k-dominant skyline of a random sample. Then

$$\mathbb{E}[M_{d,k}(n)] \to 0 \qquad (1 \leqslant k < d; d \geqslant 2),$$

as $n \to \infty$ and $d = O(1)$.

Again either zero or infinity

The phase change is more abrupt here:

$$\mathbb{E}[M_{d,k}^{[s]}(n)] = \begin{cases} O\left(n\left(1 - d^{1-d}\right)^n\right), & \textbf{if } 1 \leqslant k < d; \\ O\left(n^{1-1/d}\right), & k = d. \end{cases}$$

AN ARTIFICIAL MODEL

$\mathscr{D} := \{(-t, -2t, 3t, 4t) : 1 \leqslant t \leqslant 2\}$ $(d = 4, k = 3)$

Any two points in A are incomparable (none k-dominating the other).

Sample n points uniformly and independently in $\mathscr{D}$. Then

$$M_{d,k}(n) \sim n \qquad (\textbf{almost surely}).$$

CHANGE CONTINUOUS TO DISCRETE

A categorical model

Take n points uniformly and independently from the product space

$$\mathscr{P} := \prod_{1 \leqslant j \leqslant d} \{1, 2, \ldots, C_j\} \qquad (C_j \geqslant 2, 1 \leqslant j \leqslant d).$$

Theorem. (Categorical model)

The expected number of k-dominant skylines satisfies

$$\frac{\mathbf{E}[M_{d,k}^{[c]}(n)]}{n} \to \frac{1}{C} \qquad (1 \leqslant k \leqslant d; d \geqslant 2),$$

as $n \to \infty$, where $C := \prod_{1 \leqslant j \leqslant d} C_j$.

ASYMPTOTICALLY BINOMIAL

Theorem. (Categorical model)

Since the possible number of points is finite,

$$M_{d,k}^{[c]}(n) \sim \textbf{Binomial}\left(n; \frac{1}{C}\right).$$

Now too many candidate points!!

k-DOMINANT SKYLINE USELESS?

Why still so popular?

Gap between theory and practice

BRIDGE: UNIFORM ASYMPTOTICS

n never ∞ in practice

If $n = 10^5$, $d = 10$, then $d = O(1)$ or $O(\log n)$ or $O(n^{1/4})$?

Consider $n \to \infty$ and d varies with n

Lemma (Hypercube model, $k = d - 1$)

If $d = o\left(\frac{\log n}{\log\log n}\right)$, then

$$\mathbb{E}[M_{d,d-1}(n)] \sim \Gamma\left(\frac{1}{d-1}\right)^d \frac{n^{-1/(d-1)}}{d-1},$$

uniformly in d for large n.

$\Gamma(x) \sim x^{-1}$ **as** $x \to 0$

BRIDGE: UNIFORM ASYMPTOTICS

n never ∞ in practice

If $n = 10^5$, $d = 10$, then $d = O(1)$ or $O(\log n)$ or $O(n^{1/4})$?

Consider $n \to \infty$ and d varies with n

Lemma (Hypercube model, $k = d - 1$)

If $d = o\left(\frac{\log n}{\log\log n}\right)$, then

$$\mathbb{E}[M_{d,d-1}(n)] \sim \Gamma\left(\frac{1}{d-1}\right)^d \frac{n^{-1/(d-1)}}{d-1},$$

uniformly in d for large n.

$\Gamma(x) \sim x^{-1}$ **as** $x \to 0$

THRESHOLD PHENOMENON

Theorem. (Hypercube model, $k = d - 1$)

Let

$$d_0 = \left\lfloor \sqrt{\frac{2\log n}{W(2\log n)}} \right\rfloor + 1,$$

where W denotes the Lambert-W function ($W(x)e^{W(x)} = x$). Then the expected number of $(d-1)$-dominant skyline points satisfies

$$\mathbf{E}[M_{d,d-1}(n)] \to \begin{cases} 0, & \textbf{\textit{if}}\ d \leqslant d_0 - 1; \\ 0, & \textbf{\textit{if}}\ d = d_0\ \textbf{\textit{and}}\ n \neq i^{i^2}; \\ 1, & \textbf{\textit{if}}\ d = d_0\ \textbf{\textit{and}}\ n = i^{i^2} (i \geqslant 2); \\ \infty, & \textbf{\textit{if}}\ d \geqslant d_0. \end{cases}$$

$$W(\log x) \sim \log x - \log\log x$$

d_0 GROWS VERY SLOWLY

Let $a_i := i^{i^2}$ with $a_1 := 2$.

Then

$$d_0 = \left\lfloor \sqrt{\frac{2\log n}{W(2\log n)}} \right\rfloor + 1 = i + 1 \textbf{ for } a_i \leqslant n < a_{i+1}.$$

$$d_0 = \begin{cases} 2, & \textbf{if } 2 \leqslant n \leqslant 15; \\ 3, & \textbf{if } 16 \leqslant n \leqslant 19682; \\ 4, & \textbf{if } 19683 \leqslant n \leqslant 4294967295; \\ 5, & \textbf{if } 4294967296 \leqslant n \leqslant 2.98\cdots \times 10^{17}; \\ 6, & \textbf{if } 2.98\cdots \times 10^{17} \leqslant n \leqslant 1.03\cdots \times 10^{28}; \\ 7, & \textbf{if } 1.03\cdots \times 10^{28} \leqslant n \leqslant 2.56\cdots \times 10^{41}. \end{cases}$$

THRESHOLD PHENOMENON

Theorem. (d-simplex, $k = d - 1$)

Let

$$d_1 := \left\lfloor \frac{\log n}{W(\log n)} \right\rfloor .$$

Then

$$\mathbf{E}[M^{[s]}_{d,d-1}(n)] \to \begin{cases} 0, & \textit{if } d \leqslant d_1; \\ \infty, & \textit{if } d \geqslant d_1 + 1. \end{cases}$$

Lemma. Uniformly for $2 \leqslant d \leqslant d_1 + 3$

$$\mathbf{E}[M^{[s]}_{d,d-1}(n)] \sim \alpha_d n^{-(d-1)/2} \left(1 - d^{1-d}\right)^n ,$$

where

$$\alpha_d := \frac{(d-2)!}{\sqrt{d}} \left(\frac{2\pi(d^{d-1} - 1)}{d-1} \right)^{(d-1)/2} .$$

A SKETCH OF PROOF

$M_{d,k}(n) \leqslant M_{d,k+1}(n)$

$$\mathbb{E}[M_{d,d-1}(n)] = n \int_{[0,1]^d} (1 - |B_{d-1}(\mathbf{x})|)^{n-1}\, \mathbf{dx},$$

where $B_{d-1}(\mathbf{x})$ denotes the region of the points in $[0,1]^d$ that k-dominate $\mathbf{x}$

$$B_{d-1}(\mathbf{x}) = \sum_{1 \leqslant \ell \leqslant d} \prod_{j \neq \ell} x_j - (d-1) \prod_{1 \leqslant j \leqslant d} x_j.$$

A NUMERICAL EXAMPLE

Hypercube model, $k = d - 1$

The expected number of $(d-1)$-skyline points are given approximately by

n \ d	4	5	6	7
10^4	0.8	15.7	296.8	5945.2
10^5	0.4	8.8	187.3	4050.4

Choose $d = 5$

OPEN QUESTIONS

Asymptotic estimates for $\mathbb{E}[M_{d,k}(n)]$?

$$\mathbb{E}[M_{d,k}(n)] \asymp n^{1-d/k}$$

Uniform asymptotic approximations in d, k?

Threshold phenomenon for other k?

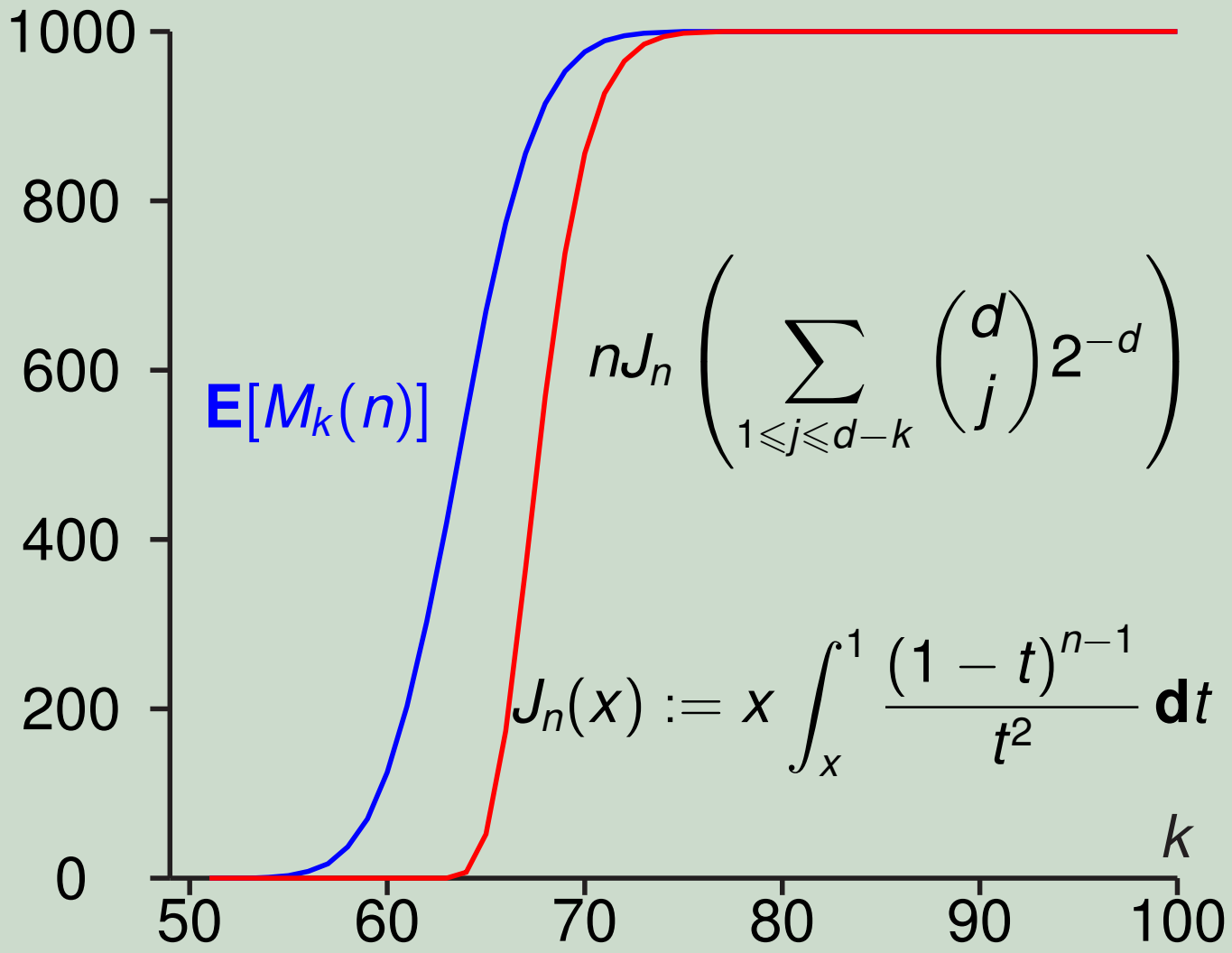
difficult in general, $k > d/2$?

Variance?

also hard

A LOWER BOUND FOR $\mathbb{E}[M_{d,k}(n)]$

$n = 1000$, $d = 100$ and k from 50 to 100



DOMINANCE, MAXIMA AND RECORDS