PKU 2014

New Problems and Techniques in Stochastic Combinatorial Optimization

Jian Li

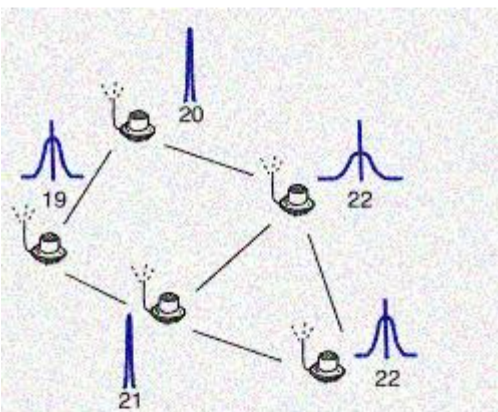
Institute of Interdisciplinary Information Sciences

Tsinghua University

lijian83@mail.tsinghua.edu.cn

Uncertain Data and Stochastic Model

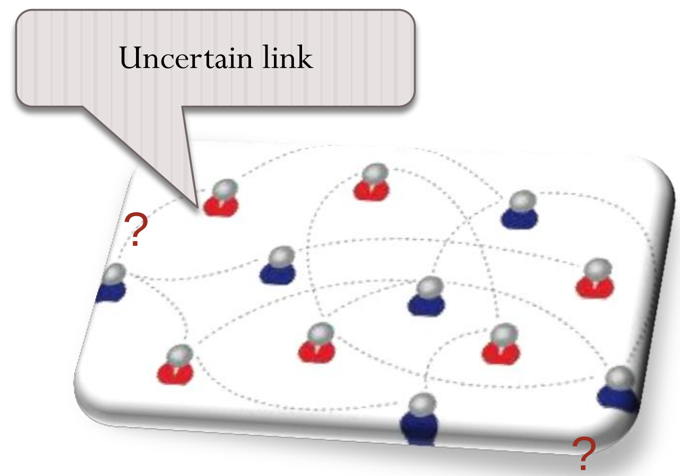
- Data Integration and Information Extraction
- Sensor Networks; Information Networks
- Probabilistic models in machine learning



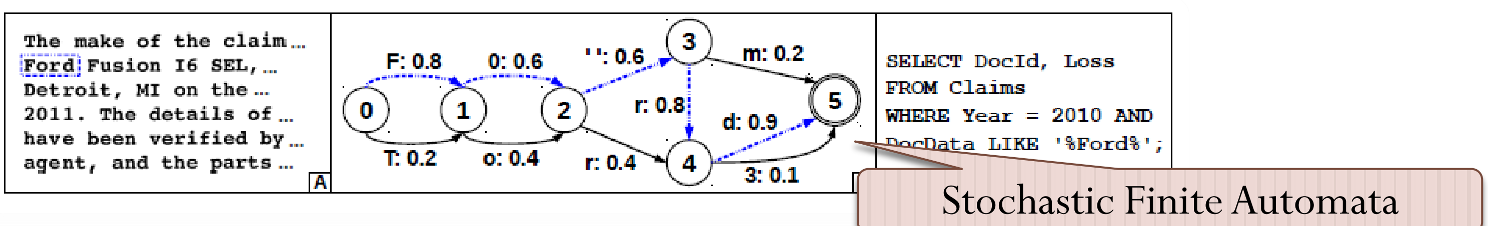
Sensor Readings

Sensor ID	Temp.
1	Gauss(40,4)
2	Gauss(50,2)
3	Gauss(20,9)
…	…

Probabilistic database



Social networks



Uncertain Data and Stochastic Model

- Future data are usually modeled by stochastic models

Dealing with Uncertainty

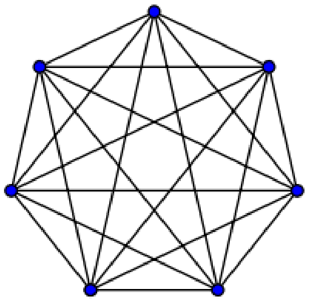
- Handling uncertainty is a very broad topic that spans multiple disciplines
  - Economics / Game Theory
  - Finance
  - Operation Research
  - Management Science
  - Probability Theory / Statistics
  - Psychology
  - Computer Science

Today: Problems in **Stochastic Combinatorial Optimization**

Outline

- A Classical Example: E[MST] in $[0,1]^2$
- Estimating E[MST] and other statistics
- Expected Utility Theory
  - Expected Utility Maximization
  - Threshold Probability Maximization
- The Poisson Approximation Technique
  - Expected Utility Maximization
  - Stochastic Knapsack
  - Other Applications
- Conclusion

Ignoring uncertainty is not the right thing to do

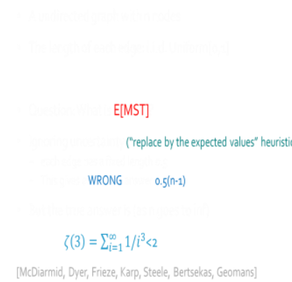


- A undirected graph with n nodes
- The length of each edge: i.i.d. Uniform[0,1]

- Question: What is E[MST]?
- Ignoring uncertainty ("replace by the expected values" heuristic)
  - each edge has a fixed length 0.5
  - This gives a WRONG answer 0.5(n-1)

Ignoring uncertainty is not the right thing to do

- A undirected graph with n nodes
- The length of each edge: i.i.d. Uniform[0,1]



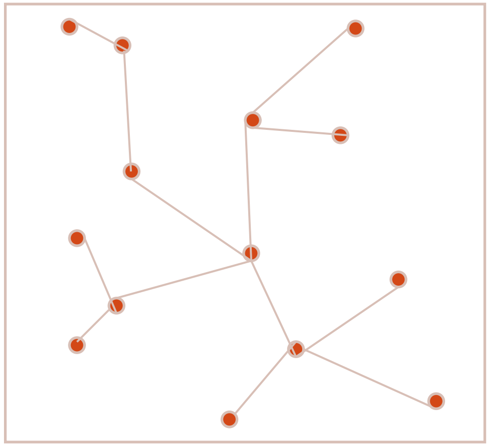
- Question: What is E[MST]?
- Ignoring uncertainty ("replace by the expected values" heuristic)
  - each edge has a fixed length 0.5
  - This gives a WRONG answer 0.5(n-1)
- But the true answer is (as n goes to inf)

$$\zeta(3) = \sum_{i=1}^{\infty} 1/i^3 < 2$$

[McDiarmid, Dyer, Frieze, Karp, Steele, Bertsekas, Geomans]

A Similar Problem

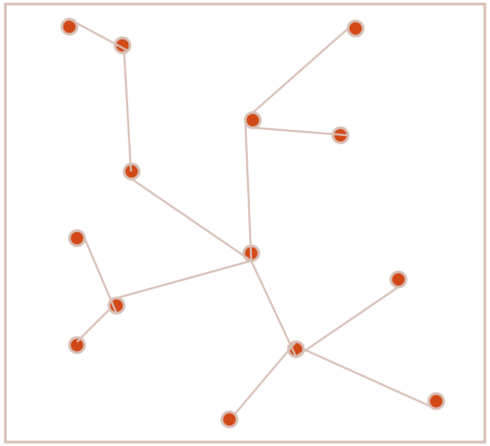
- N points: i.i.d. uniform$[0,1] \times [0,1]$



- Question: What is E[MST] ?
- Answer:

A Similar Problem

- N points: i.i.d. uniform$[0,1] \times [0,1]$



- Question: What is E[MST] ?
- Answer: $\theta(\sqrt{n})$ [Frieze, Karp, Steele, …]

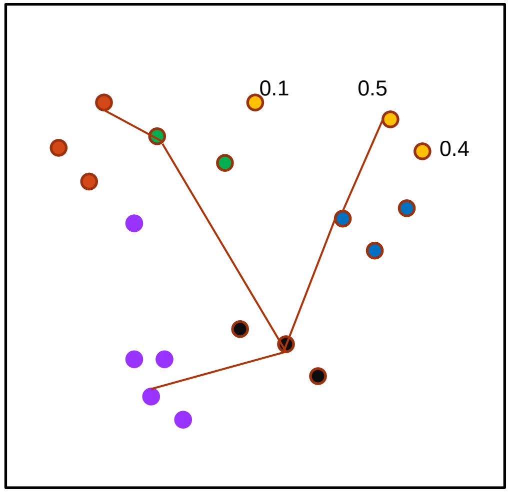
The problem is similar, but the answer is not similar…………

Outline

- A Classical Example: E[MST] in $[0,1]^2$
- Estimating E[MST] and other statistics
- Expected Utility Theory
  - Expected Utility Maximization
  - Threshold Probability Maximization
- The Fourier Series Technique
  - Expected Utility Maximization
- The Poisson Approximation Technique
  - Expected Utility Maximization
  - Stochastic Knapsack
  - Other Applications
- Conclusion

A Computational Problem

- The position of each point is random (non-i.i.d)
- A model in wireless networks



- Question: What is E[MST] ?
- Of Course, there is no close-form formula
- Need efficient algorithms to compute E[MST]

[Huang, L. ArXiv 2013]

Our Results

Problems		Existential	Locational
Closest Pair (Section 2)	$\mathbb{E}[\mathsf{CP}]$	#P/FPRAS	#P/FPRAS
	$\Pr[\mathsf{CP} \le 1]$	#P [19]/FPRAS	#P/FPRAS
	$\Pr[\mathsf{CP} \ge 1]$	#P/Inapprox	#P/Inapprox
Diameter (Section 2.2)	$\mathbb{E}[\mathsf{D}]$	#P/FPRAS	#P/FPRAS
	$\Pr[\mathsf{D} \le 1]$	#P/Inapprox	#P/Inapprox
	$\Pr[\mathsf{D} \ge 1]$	#P/FPRAS	#P/FPRAS
MST (Section 3)	$\mathbb{E}[\mathsf{MST}]$	#P [20]/FPRAS	#P [20]/FPRAS
	$\Pr[\mathsf{MST} \le 1]$	#P/Inapprox [20]	#P/Inapprox [20]
	$\Pr[\mathsf{MST} \ge 1]$	#P/Open	#P/Open
k-Clustering/k-Center/k-median (Section 4)	$\mathbb{E}[\mathsf{C}_k]$	#P/FPRAS*1	#P/FPRAS*
Perfect Matching (Section 5)	$\mathbb{E}[\mathsf{MPM}]$	\	Open/FPRAS
Cycle Cover (Section 6)	$\mathbb{E}[\mathsf{CC}]$	#P/FPRAS	#P/FPRAS
kth Closest Pair (Section 4)	$\mathbb{E}[\mathsf{CP}_k]$	#P [19]/FPRAS	#P/Open
kth Longest m-Nearest Neighbor	$\mathbb{E}[k\text{-}\mathsf{NN}_m]$	#P/Open	#P/Open
Convex Hull (2D) (Section 7)	$\mathbb{E}[\mathsf{CH}]$	Open/FPRAS	Open

MST over Stochastic Points

- The problem is #P-hard [Kamousi, Chan, Suri. SoCG'11]
- So, let us focus on approximating the value
- Attempt one: list all realizations? (Exponentially many)
- Attempt two: Monte Carlo (variance can be very large)

PDF of a random var

A sufficient condition for MC to work (in poly time): $$\frac{Max\{X\}}{\mathrm{E}[X]} \leq \text{poly}$$

(just Chernoff Bound)

MST over Stochastic Points

- Our approach: (sketch)
- Law of total expectation:

$$\mathbf{E}[X] = \sum_{y} \Pr[Y = y]\, \mathbf{E}[X \mid Y = y]$$

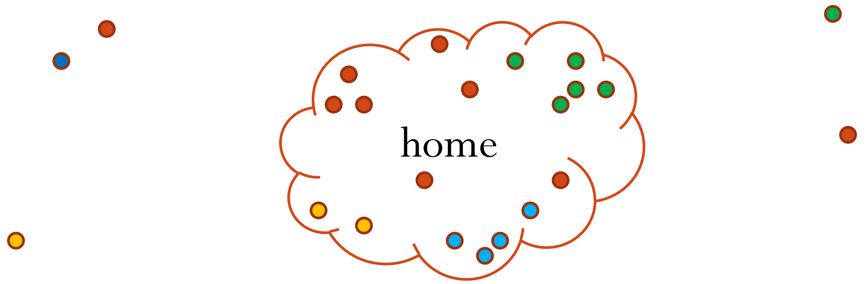
A carefully chosen random event Y

Hopefully, we have Easy to compute Low variance

How to choose Y?

MST over Stochastic Points

- The "home set" technique:



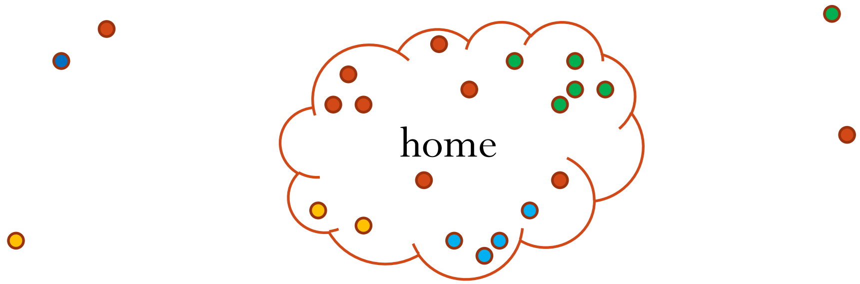
(1) $\Pr[\mathit{all\ nodes\ are\ at\ home}] \approx 1$

(2) $\mathbf{E}[\text{MST} \mid \mathit{all\ node\ are\ at\ home}]$ can be estimated:

$$\frac{\text{Diameter(home)}}{\text{E[MST} \mid \text{ all node are at home]}} \leq \text{poly}$$

MST over Stochastic Points

- The "home set" technique:



(1) $\Pr[all\ nodes\ are\ at\ home] \approx 1$

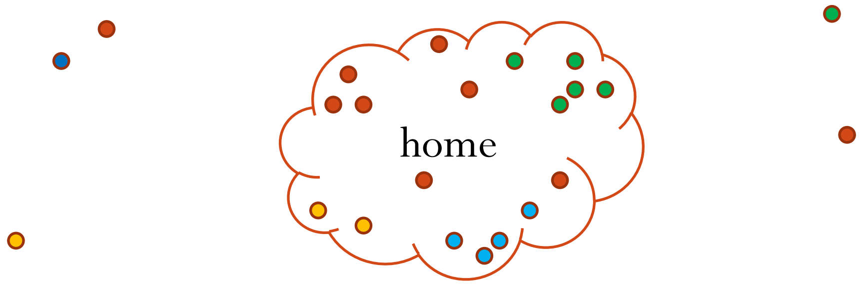
(2) $\mathbf{E}[\text{MST} \mid all\ node\ are\ at\ home]$ can be estimated (due to low variance)

$$\frac{\text{Diameter(home)}}{\text{E[MST} \mid \text{ all node are at home]}} \leq \text{poly}$$

Home={all points w.p. $\geq 1/(nm)^2$}

MST over Stochastic Points

- The "home set" technique:



(1) $\Pr[\textit{all nodes are at home}] \approx 1$

(2) $\mathbf{E}[\text{MST} \mid \textit{all node are at home}]$ can be estimated (due to low variance)

(3)

$$\mathbf{E}[MST] = \sum_y \Pr[y \textit{ nodes are at home}]\, \mathbf{E}[X \mid y \textit{ nodes are at home}]$$

$$\approx \Pr[\textit{all nodes are at home}]\, \mathbf{E}[X \mid \textit{all nodes are at home}] + \Pr[n-1 \textit{ nodes are at home}]\, \mathbf{E}[X \mid n-1 \textit{ nodes are at home}]$$

The contribution of other terms is negligible and can be ignored.

Estimating Statistics

- Another technique based on Hierarchical tree decomposition
- Interesting connection to classical counting problem:
  - Counting #perfect matchings
  - Counting #Knapsack
  - Counting #(certain subgraphs)
- Still some open questions

Outline

- A Classical Example: E[MST] in $[0,1]^2$
- Estimating E[MST] and other statistics
- Expected Utility Theory
  - Expected Utility Maximization
  - Threshold Probability Maximization
- The Poisson Approximation Technique
  - Expected Utility Maximization
  - Stochastic Knapsack
  - Other Applications
- Conclusion

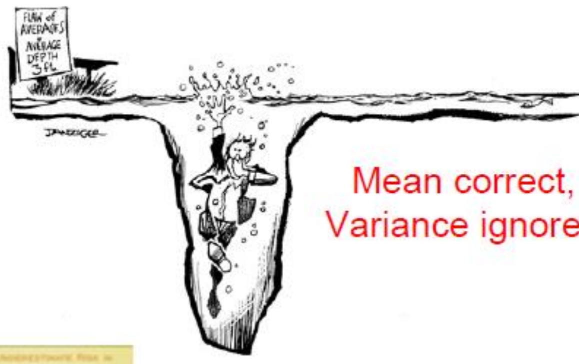
Inadequacy of Expected Value

- Stochastic Optimization
  - Some part of the input are probabilistic
  - Most common objective: Optimizing the expected value

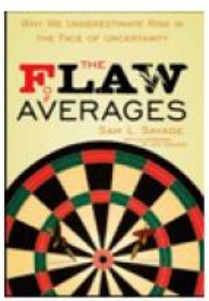
Inadequacy of Expected Value

- Be aware of risk!

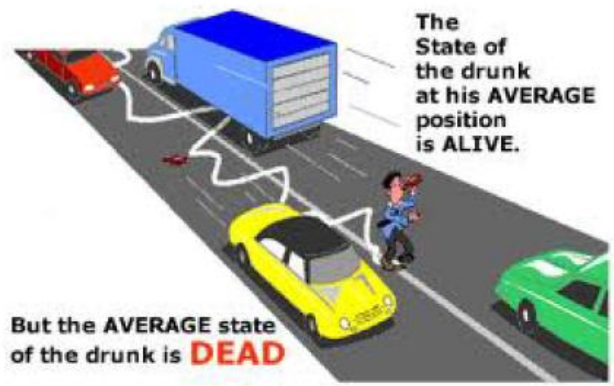
Flaw of averages (weak form):



Mean correct,
Variance ignored



Flaw of averages (strong form):



Wrong value of mean:
$f(E[X]) \neq E[f(X)]$

21

- St. Petersburg Paradox

Inadequacy of Expected Value

- Inadequacy of expected value:
  - Unable to capture risk-averse or risk-prone behaviors
    - Action 1: $100 VS Action 2: $200 w.p. 0.5; $0 w.p. 0.5
    - Risk-averse players prefer Action 1
    - Risk-prone players prefer Action 2 (e.g., a gambler spends $100 to play Double-or-Nothing)
- St. Petersburg Paradox
  - You pay x dollars to enter the game
    - Repeatedly toss a fair coin until a tail appears
    - payoff=2^k where k=#heads
  - How much should x be?
    - Expected payoff =
    - Few people would pay even $25 [Martin '04]

Expected Utility Maximization

Remedy: Use a utility function

$\mu : R \to R$: The utility function: value (profit/cost)-> utility

Expected Utility Maximization Principle: the decision maker should choose the action that maximizes the **expected utility**

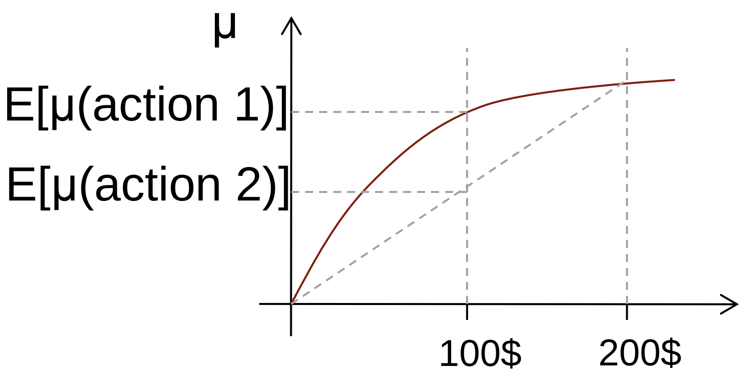
$$\text{maximize. } \mathbb{E}[\mu(\text{profit})]$$

- Proved quite useful to explain some popular choices that seem to contradict the expected value criterion

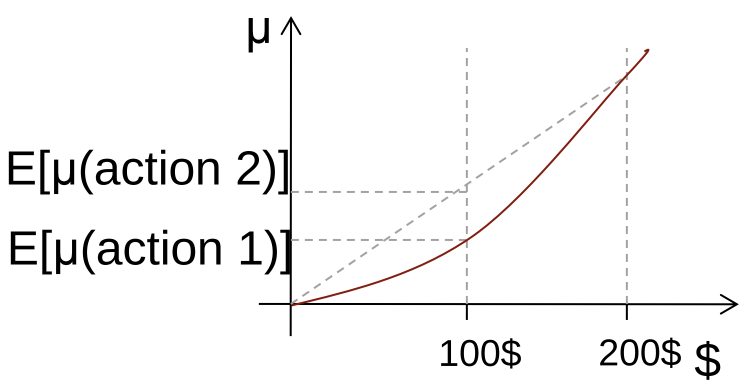
Expected Utility Maximization Principle

Expected Utility Maximization Principle: the decision maker should choose the action that maximizes the **expected utility**

- Action 1: \$100
- Action 2: \$200 w.p. 0.5; \$0 w.p. 0.5



Risk-averse

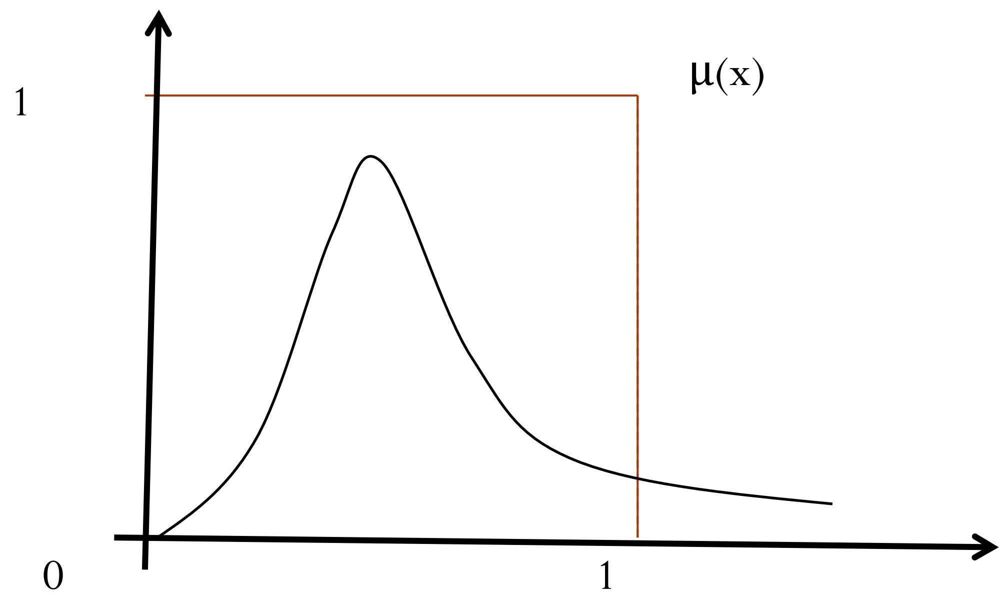


Risk-prone

- Von Neumann and Morgenstern provides an *axiomitization* of the principle (known as von Neumann-Morgenstern expected utility theorem).

Threshold Probability Maximization

- If μ *is a threshold function, maximizing* *E[μ(cost)]* *is equivalent to maximizing* ***Pr[w(cost)<1]***
  - *minimizing overflow prob.* [Kleinberg, Rabani, Tardos. STOC'97] [Goel, Indyk. FOCS'99]
  - *chance-constrained stochastic optimization problem* [Swamy. SODA'11]



Outline

- A Classical Example: E[MST] in $[0,1]^2$
- Estimating E[MST] and other statistics
- Expected Utility Theory
  - Expected Utility Maximization
  - Threshold Probability Maximization
- The Poisson Approximation Technique
  - Expected Utility Maximization
  - Stochastic Knapsack
  - Other Applications
- Conclusion

New Techniques

- A common challenge: How to deal with/ optimize on the distribution of the sum of several random variables.
  - More often seen in the risk-aware setting (linearity of expectation does not help)
- Previous techniques:
  - Special distributions [Nikolova, Kelner, Brand, Mitzenmacher. ESA'06] [Goel, Indyk. FOCS'99] [Goyal, Ravi. ORL09] …..
  - Effective bandwidth [Kleinberg, Rabani, Tardos STOC'97]
  - LP [Dean, Goemans, Vondrak. FOCS'04] …..
  - Discretization [Bhalgat, Goel, Khanna. SODA'11]
  - Characteristic Function + Fourier Series Decomposition [L, Deshpande. FOCS'11]
- Today: Poisson Approximation [L, Yuan STOC'13]

Threshold Probability Maximization

- Deterministic version:
  - A set of element $\{e_i\}$, each associated with a weight w_i
  - A solution S is a subset of elements (that satisfies some property)
  - **Goal:** Find a solution S such that the total weight of the solution $w(S)=\Sigma_{i\in S}w_i$ is minimized
  - E.g. shortest path, minimal spanning tree, top-k query, matroid base

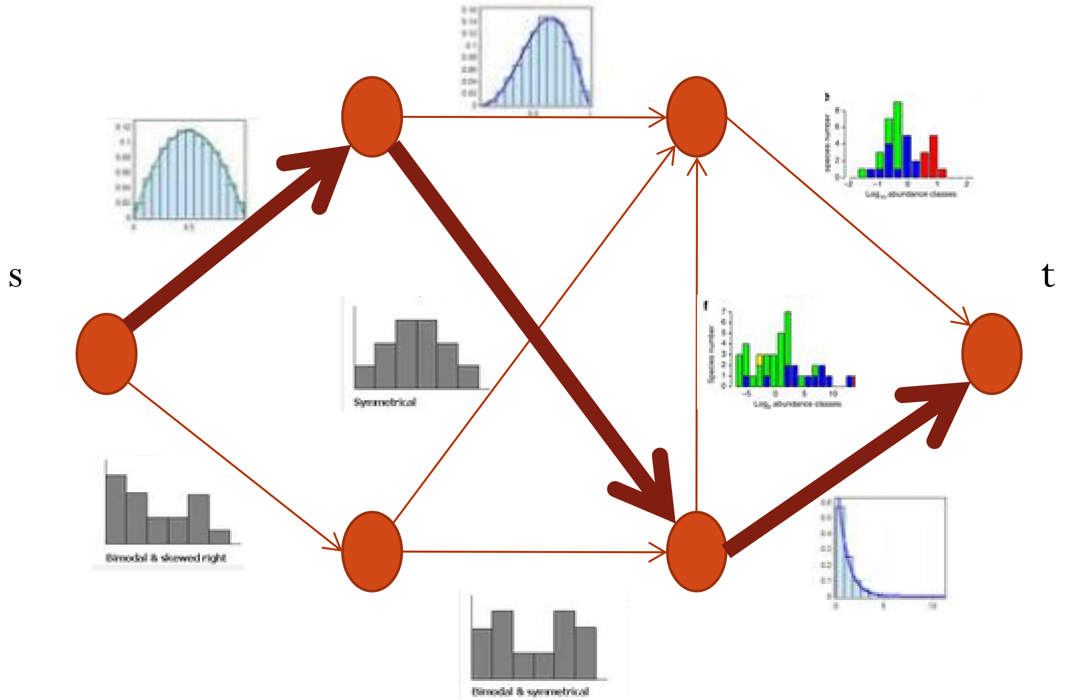
Threshold Probability Maximization

- Deterministic version:
  - A set of element $\{e_i\}$, each associated with a weight w_i
  - A solution S is a subset of elements (that satisfies some property)
  - **Goal:** Find a solution S such that the total weight of the solution $w(S)=\Sigma_{i\epsilon S} w_i$ is minimized
  - E.g. shortest path, minimal spanning tree, top-k query, matroid base
- Stochastic version:
  - w_is are independent positive random variables
  - **Goal:** Find a solution S such that the *threshold probability*

$$\Pr[w(S) \leq 1] \quad \text{is maximized.}$$

Threshold Probability Maximization

- **Stochastic shortest path** : find an s-t path P such that $Pr[w(P) < 1]$ is maximized



Our Result

If the deterministic problem is "easy", then for any $\epsilon > 0$, we can find a solution S such that

$$\Pr[w(S) \leq 1 + \epsilon] > OPT - \epsilon$$

"Easy": there is a PTAS for the corresponding O(1)-dim packing problem:

- Shortest path, MST, matroid base, matroid intersection, min-cut

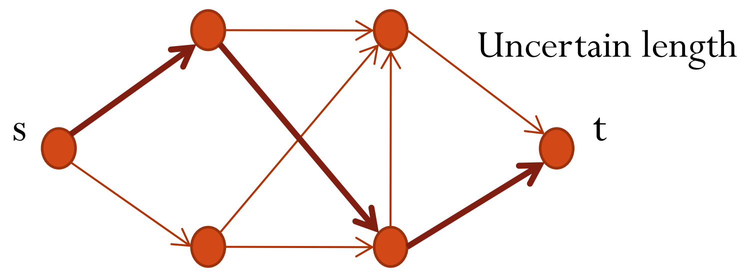
The above result can be generalized to the expected utility maximization problem:

maximize $\mathrm{E}[\mu(X(S))]$ for Lipschitz utility function μ

- generalizes/simplies/improves the previous results in [Nikolova, Kelner, Brand, Mitzenmacher. ESA'06] [Nikolova. APPROX'10] [Kleinberg, Rabani, Tardos. STOC'97] [Goel, Indyk. FOCS'99] [Goyal, Ravi. ORL09] [Bhalgat, Goel, Khanna. SODA'11] [Li, Deshpande. FOCS'11]

Our Results

- **Stochastic shortest path** : find an s-t path P such that $Pr[w(P)<1]$ is maximized



- Previous results
  - Many heuristics
  - Poly-time approximation scheme (PTAS) if (1) all edge weights are normally distributed r.v.s (2) $OPT>0.5$[Nikolova, Kelner, Brand, Mitzenmacher. ESA'06] [Nikolova. APPROX'10]
  - Bicriterion PTAS ($Pr[w(P)<1+\delta]>(1\text{-eps})OPT$) for exponential distributions [Nikolova, Kelner, Brand, Mitzenmacher. ESA'06]
- Our result
  - Bicriterion PTAS if $OPT= Const$

Our Results

- **Stochastic knapsack**: find a collection S of items such that $Pr[w(S)<1]>\gamma$ and the total profit is maximized

Each item has a deterministic profit and a (uncertain) size

Knapsack, capacity=1

- Previous results
  - $log(1/(1-\gamma))$-approximation [Kleinberg, Rabani, Tardos. STOC'97]
  - Bicriterion PTAS for exponential distributions [Goel, Indyk. FOCS'99]
  - PTAS for Bernouli distributions if $\gamma = Const$ [Goel, Indyk. FOCS'99] [Chekuri, Khanna. SODA'00]
  - Bicriterion PTAS if $\gamma = Const$ [Bhalgat, Goel, Khanna. SODA'11]
- Our result
  - Bicriterion PTAS if $\gamma = Const$ (with a better running time than Bhalgat et al.)
  - Stochastic partial-ordered knapsack problem with tree constraints

Outline

- A Classical Example: E[MST] in $[0,1]^2$
- Estimating E[MST] and other statistics
- Expected Utility Theory
  - Expected Utility Maximization
  - Threshold Probability Maximization
- The Fourier Series Technique
  - Expected Utility Maximization
- The Poisson Approximation Technique
  - Expected Utility Maximization
  - Stochastic Knapsack
  - Other Applications
- Conclusion

Algorithm2 (based on Poisson Approx)

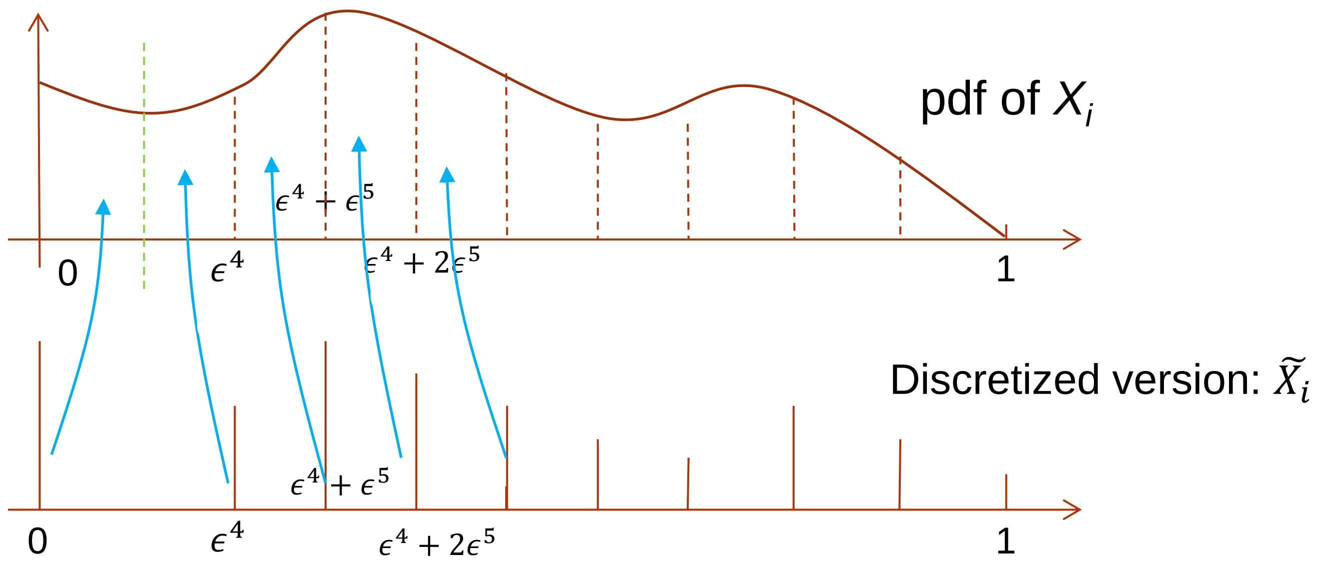
- Step 1: Discretizing the prob distr

  (Similar to [Bhalgat, Goel, Khanna. SODA'11], but much simpler)
- Step 2: Reducing the problem to the multi-dim problem

Algorithm2 (based on Poisson Approx)

- Step 1: Discretizing the prob distr

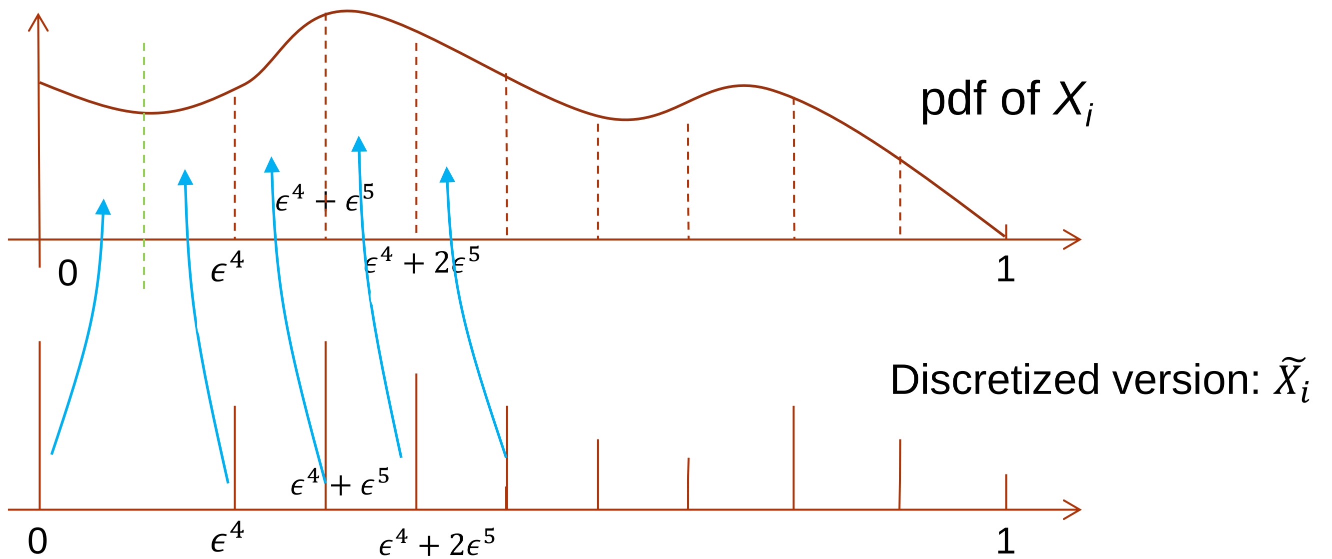
(Similar to [Bhalgat, Goel, Khanna. SODA'11], but simpler)



Algorithm2 (based on Poisson Approx)

- Step 1: Discretizing the prob distr

(Similar to [Bhalgat, Goel, Khanna. SODA'11], but simpler)



The behaviors of $\widetilde{X}_i$ and X_i are close:

1. $\Pr[X(S) \le \beta] \le \Pr[\widetilde{X}(S) \le \beta + \epsilon] + O(\epsilon)$;
2. $\Pr[\widetilde{X}(S) \le \beta] \le \Pr[X(S) \le \beta + \epsilon] + O(\epsilon)$.

Algorithm2 (based on Poisson Approx)

- Step 2: Reducing the problem to the multi-dim problem
  - Heavy items: $E[X_i]>poly(\epsilon)$
    - At most $O(1/poly(\epsilon))$ many heavy items, so we can afford enumerating them

Algorithm2 (based on Poisson Approx)

- Step 2: Reducing the problem to the multi-dim problem
  - Heavy items: $E[X_i]>poly(\epsilon)$
    - At most $O(1/poly(\epsilon))$ heavy items, so we can afford enumerating them
  - Light items:
    - Fix the set H of heavy items
    - Each X_i can be represented as a O(1)-dim vector **Sg(i)** (signature)

$$\mathbf{Sg}(i) = (\Pr[\tilde{X}_i = \epsilon^4], \Pr[\tilde{X}_i = \epsilon^4 + \epsilon^5], \ldots \ldots)$$

    - Enumerating all O(1)-dim (budget) vectors B
      - Find a set S such that $S \cup H$ is feasible and

$$\mathbf{Sg}(S) = \sum_{i \in S} \mathbf{Sg}(i) \leq (1+\epsilon)B \quad \text{(using the multi-dim PTAS)}$$

(or declare there is none S s.t. $\mathbf{Sg}(S) \leq B$)

    - Return $S \cup H$ for which $\Pr[w(S \cup H) \leq 1+\epsilon]$ is largest

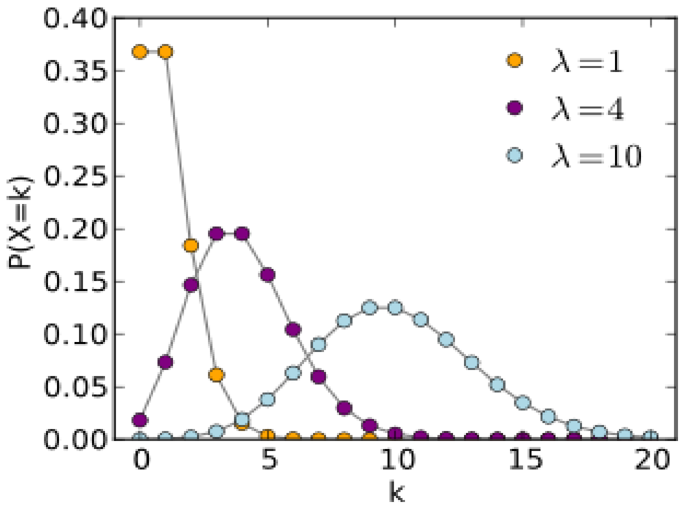
Poisson Approximation

Well known: Law of small numbers

n Bernoulli r.v. X_i (1-p, p)

$np = const$

As $n \to \infty, \sum X_i \sim Poisson(np)$



Poisson Approximation

Le Cam's theorem (rephrased):

n r.v. X_i (with common support (0,1,2,3,4,…)) with signature

$$\mathbf{sg}_i = (\Pr[X_i = 1], \Pr[X_i = 2], \ldots)$$

Let $\mathbf{sg} = \sum_i \mathbf{sg}$

Y_i are i.i.d. r.v. with distr $\mathbf{sg}/|\mathbf{sg}|_1$

Y follows the ***compound Poisson distr (CPD)*** *corresponding to* **sg**

$$Y = \sum_{i=1}^{N} Y_i \text{ where } N \sim \text{Poisson}(|\mathbf{sg}|_1)$$

Then, $\Delta(\sum X_i, Y) \le \sum p_i^2$ where $p_i = \Pr[X_i \neq 0]$

Variational distance:
$\Delta(X,Y) = \sum_i |\Pr[X = i] - \Pr[Y = i]|$

Poisson Approximation

- **Le Cam's theorem:** $\Delta(\sum X_i, Y) \leq \sum p_i^2$
- Ob: If S_1 and S_2 have the same signature, then they correspond to the same CPD
- So if $\sum_{i \in S_1} p_i^2$ and $\sum_{i \in S_2} p_i^2$ are sufficiently small, the distributions of $X(S_1)$ and $X(S_2)$ are close
- Therefore, enumerating the signature of light items suffices (instead of enumerating subsets)

Summary

- The #dimension needs to be $L = poly(1/\epsilon)$
- We solve an $poly\left(\frac{1}{\epsilon}\right)$-dim optimization problem
- The overall running time is $n^{poly(1/\epsilon)}$
- This improves the $n^{2^{poly(1/\epsilon)}}$ running time in [Bhalgat, Goel, Khanna. SODA'11]
- Can be easily extended to the multi-dimensional case, other combinatorial constraints etc.

Outline

- A Classical Example: E[MST] in $[0,1]^2$
- Estimating E[MST] and other statistics
- Expected Utility Theory
  - Expected Utility Maximization
  - Threshold Probability Maximization
- The Poisson Approximation Technique
  - Expected Utility Maximization
  - Stochastic Knapsack
  - Other Applications
- Conclusion

Stochastic Knapsack

- A knapsack of capacity C
- A set of items, each having a fixed profit
- Known: Prior distr of size of each item.
- Each time we choose an item and place it in the knapsack irrevocably
- The actual size of the item becomes known after the decision
- Knapsack constraint: The total size of accepted items <= C
- Goal: maximize E[Profit]

[L, Yuan STOC13]

Stochastic Knapsack

Previous work

- 5-approx [Dean, Goemans, Vondrak. FOCS'04]
- 3-approx [Dean, Goemans, Vondrak. MOR'08]
- $(1+\epsilon, 1+\epsilon)$-approx [Bhalgat, Goel, Khanna. SODA'11]
- 2-approx [Bhalgat 12]
- 8-approx (size&profit correlation, cancellation)

  [Gupta, Krishnaswamy, Molinaro, Ravi. FOCS'11]

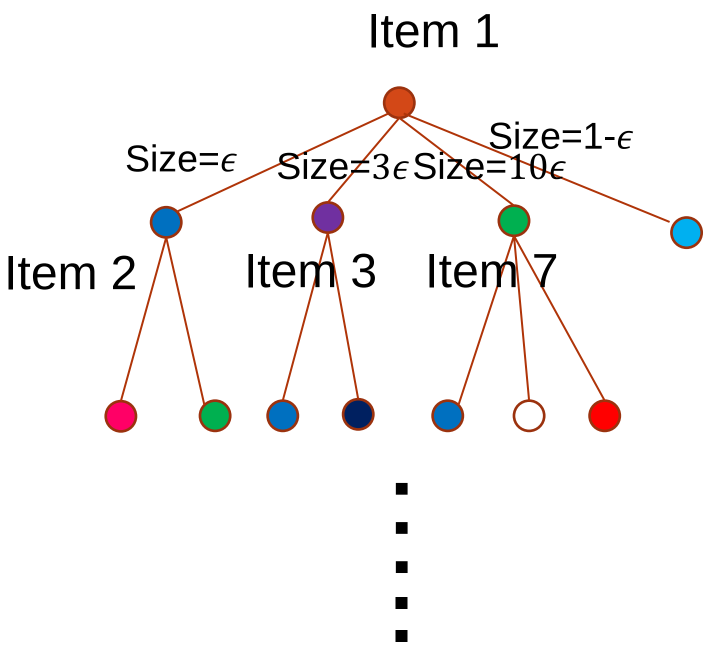
Our result:

$(1+\epsilon, 1+\epsilon)$-approx (size&profit correlation, cancellation)

2-approx (size&profit correlation, cancellation)

Stochastic Knapsack

- Decision Tree



Exponential size!!!! (depth=n)

How to represent such a tree? Compact solution?

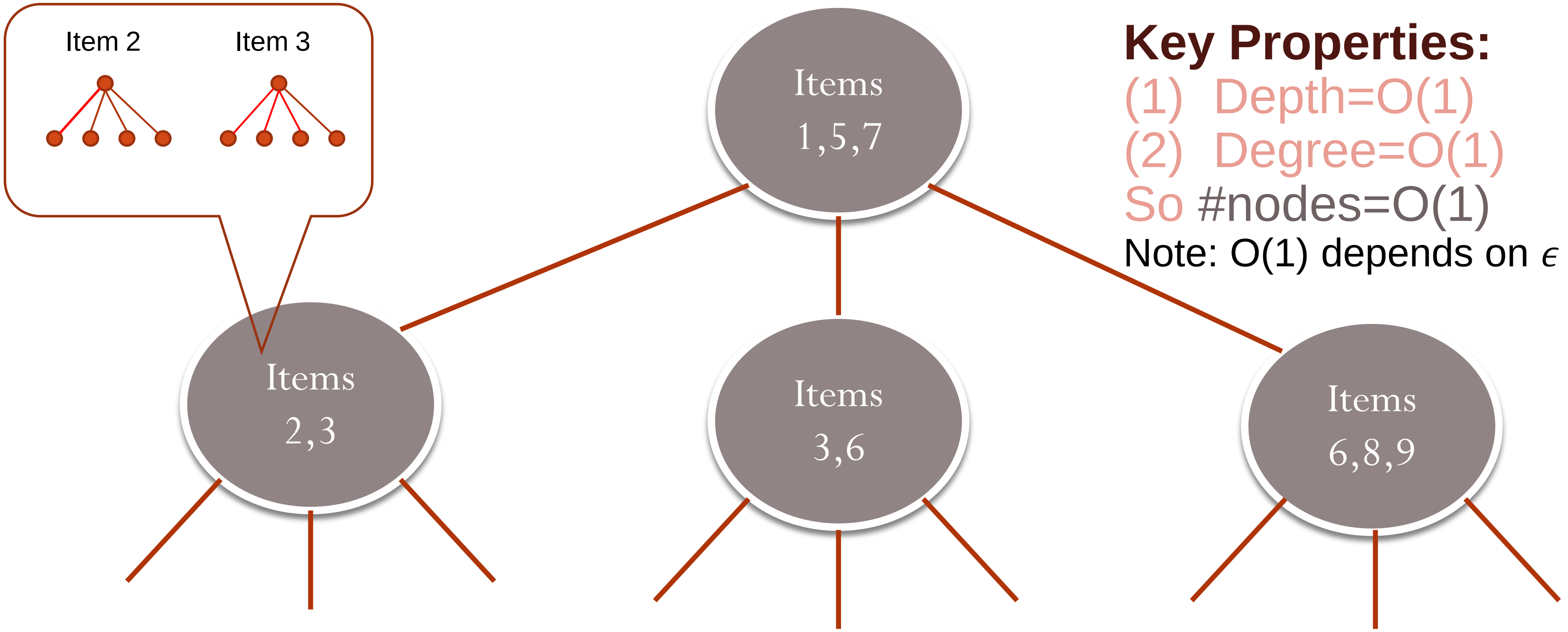
Stochastic Knapsack

- By discretization, we make some simplifying assumptions:
  - Support of the size distribution: $(0, \epsilon, 2\epsilon, 3\epsilon, \ldots \ldots, 1)$.

Still way too many possibilities, how to narrow the search space?

Block Adaptive Policies

- Block Adaptive Policies: Process items block by block

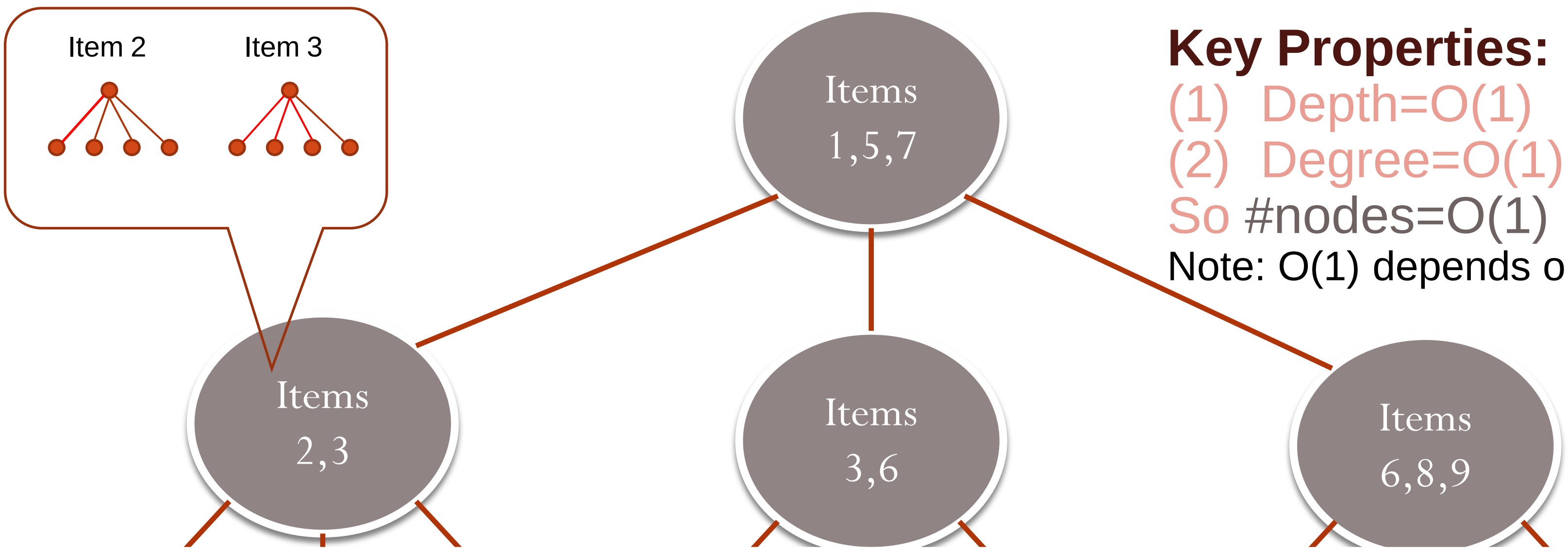


Key Properties:
(1) Depth=O(1)
(2) Degree=O(1)
So #nodes=O(1)
Note: O(1) depends on ϵ

LEMMA: [Bhalgat, Goel, Khanna. SODA'11] There is a block adaptive policy that is nearly optimal (under capacity $(1+\epsilon)C$)

Block Adaptive Policies

- Block Adaptive Policies: Process items block by block



Key Properties:
(1) Depth=O(1)
(2) Degree=O(1)
So #nodes=O(1)
Note: O(1) depends on ϵ

Still exponential many possibilities, even in a single block

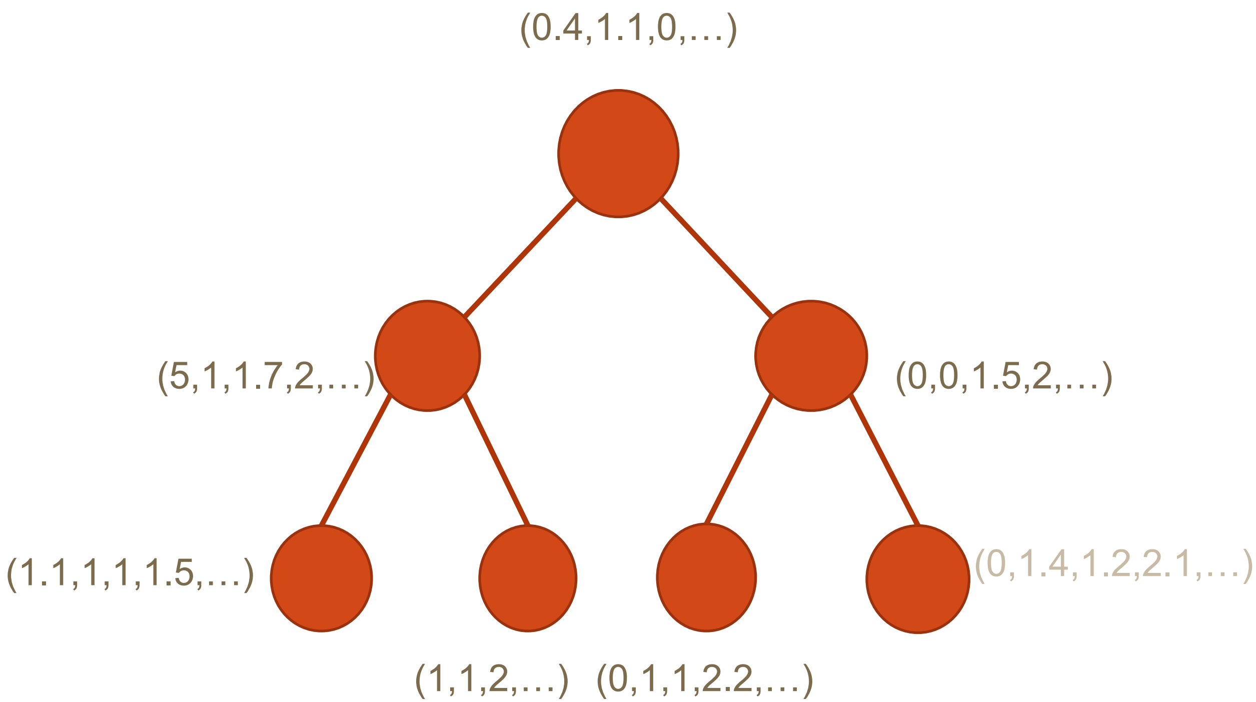
LEMMA: [Bhalgat, Goel, Khanna. SODA'11] There is a block adaptive policy that is nearly optimal (under capacity $(1+\epsilon)C$)

Poisson Approximation

- Each heavy item consists of a singleton block
- Light items:
  - Recall if two blocks have the same signature, their size distributions are similar
  - So, enumerate Signatures! (instead of enumerating subsets)

Algorithm

- Outline: Enumerate all block structures with a signature associated with each node



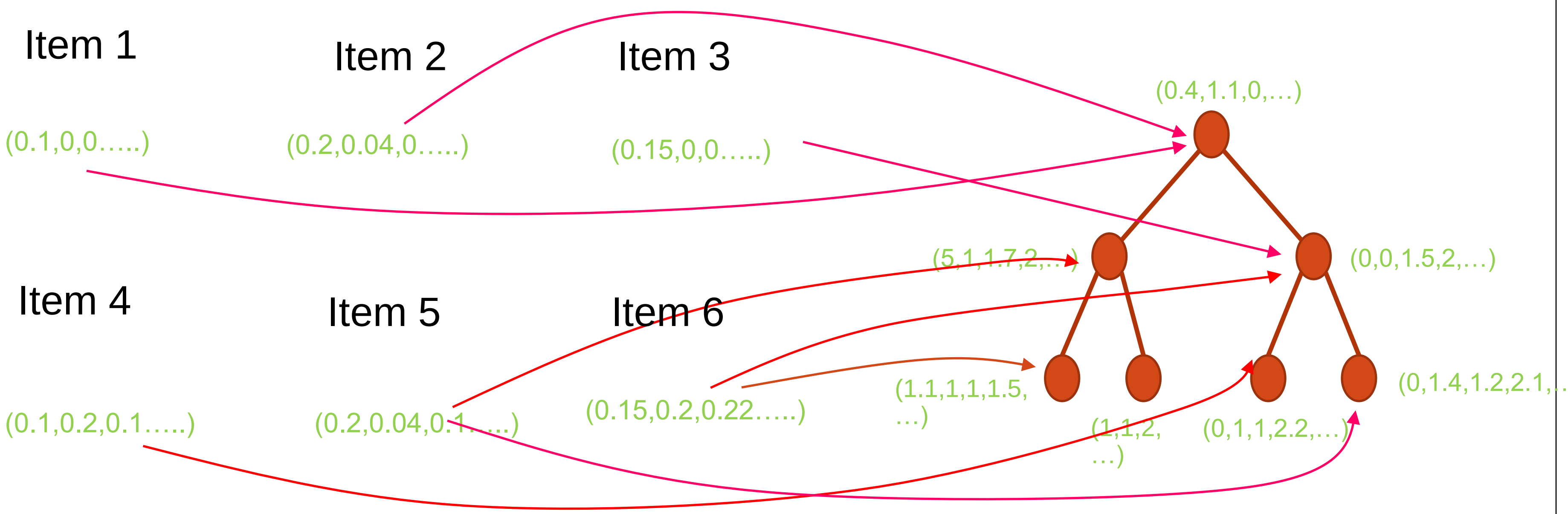
- O(1) nodes
- Poly(n) possible signatures for each node
- So total #configuration =poly(n)

Algorithm

2. Find an assignment of items to blocks that matches all signatures
   – (this can be done by standard dynamic program)

Algorithm

2. Find an assignment of items to blocks that matches all signatures
   – (this can be done by standard dynamic programming)



On any root-leaf path, each item appears at most once

Outline

- A Classical Example: E[MST] in $[0,1]^2$
- Estimating E[MST] and other statistics
- Expected Utility Theory
  - Expected Utility Maximization
  - Threshold Probability Maximization
- The Poisson Approximation Technique
  - Expected Utility Maximization
  - Stochastic Knapsack
  - Other Applications
- Conclusion

Poisson Approximation-Other Applications

- Incorporating other constraints
  - Size/profit correlation
  - cancellation
- Bayesian Online Selection Problem with Knapsack Constraint
  - Can see the actually size and profit of an item before the decision
  - $(1+\epsilon, 1+\epsilon)$-approx (against the optimal adaptive policy)

✓ Prophet inequalities [Chawla, Hartline, Malec, Sivan. STOC10] [Kleinberg, Weinberg. STOC12]
✓ Close relations with Secretary problems
✓ Applications in multi-parameter mechanism design

- Stochastic Bin Packing

[L, Yuan STOC13]

Conclusion

- Replacing the input random variable with its expectation typically is NOT the right thing to do
  - Carry the randomness along the way and optimize the expectation of the objective
- Optimizing the expectation may not be the right thing to do neither
  - Be aware of the risk
- We can often reduce the stochastic optimization problem (with independent random variables) to a constant dimensional packing problem
- Stochastic optimization problems with dependent random variables are typically extremely hard (i.e., inapproximable)

Thanks

lijian83@mail.tsinghua.edu.cn