Stochastic Combinatorial Optimization via Poisson Approximation

Jian Li, Wen Yuan
Institute of Interdisiplinary Information Sciences
Tsinghua University

STOC 2013

清華大學
Tsinghua University

lijian83@mail.tsinghua.edu.cn

Outline

- Threshold Probability Maximization
- Stochastic Knapsack
- Other Results

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

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*

$$\Pr[w(S) \leq 1]$$ is maximized.

Related Work

Studied extensively before:

- Many heuristics
- Stochastic shortest path [Nikolova, Kelner, Brand, Mitzenmacher. ESA'06] [Nikolova. APPROX'10]
- Fixed set stochastic knapsack [Kleinberg, Rabani, Tardos. STOC'97] [Goel, Indyk. FOCS'99] [Goyal, Ravi. ORL09][Bhalgat, Goel, Khanna. SODA'11]
- .....
- Chance-constrained (risk-averse) stochastic optimization problem [Swamy. SODA'11]

Related Work

Studied extensively before:

- Many heuristics
- Stochastic shortest path [Nikolova, Kelner, Brand, Mitzenmacher. ESA'06] [Nikolova. APPROX'10]
- Fixed set stochastic knapsack [Kleinberg, Rabani, Tardos. STOC'97] [Goel, Indyk. FOCS'99] [Goyal, Ravi. ORL09][Bhalgat, Goel, Khanna. SODA'11]
- .....
- Chance-constrained (risk-averse) stochastic optimization problem [Swamy. SODA'11]

A common challenge: How to deal with/ optimize on the distribution of the sum of several random variables.

Previous techniques:

- LP [Dean, Goemans, Vondrak. FOCS'04]
- Discretization [Bhalgat, Goel, Khanna. SODA'11],
- Characteristic function [Li, Deshpande. FOCS'11]

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 multi-dimensional version of the problem: Shortest path, MST, matroid base, matroid intesection, min-cut (strictly generalizing the result in [Li, Deshpande. FOCS'11])

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

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

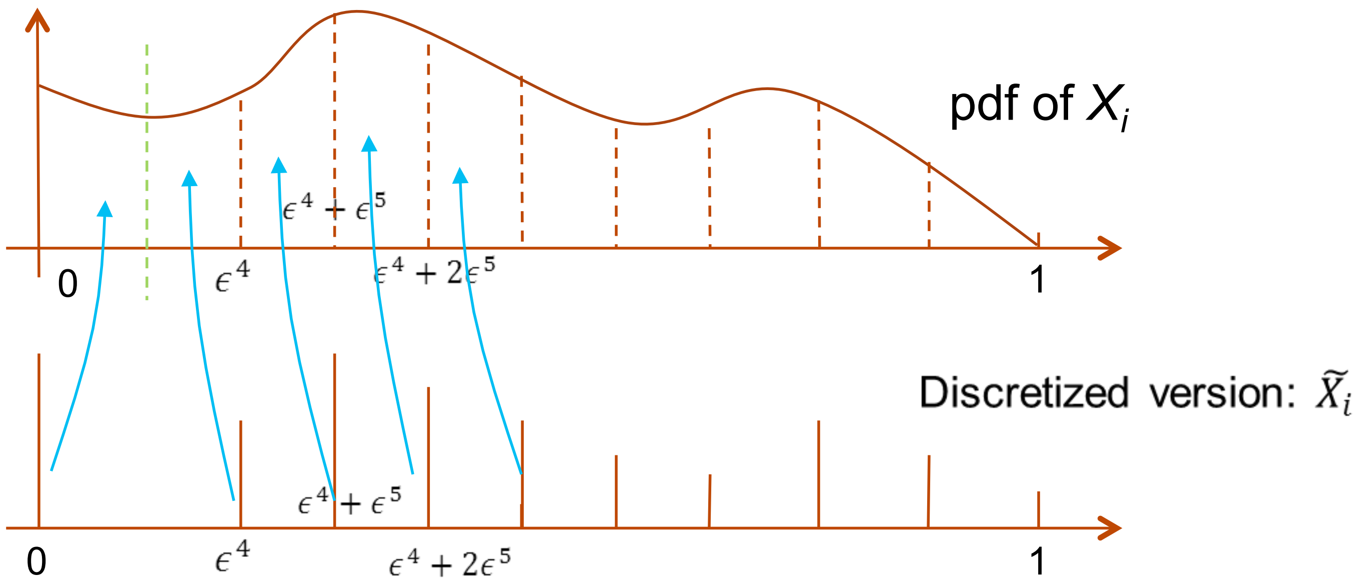
Our Algorithm

- 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

Our Algorithm

- Step 1: Discretizing the prob distr

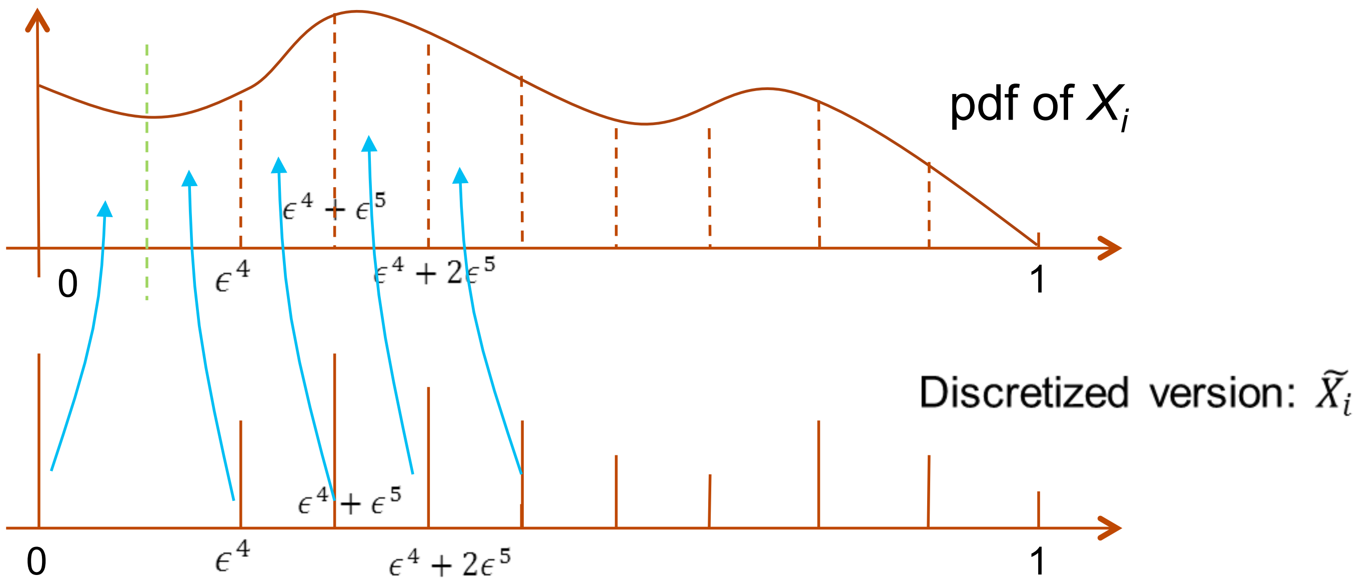
(Similar to [Bhalgat, Goel, Khanna. SODA'11], but simpler)



Our Algorithm

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

Our Algorithm

- 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

Our Algorithm

- 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:
    - 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

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

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

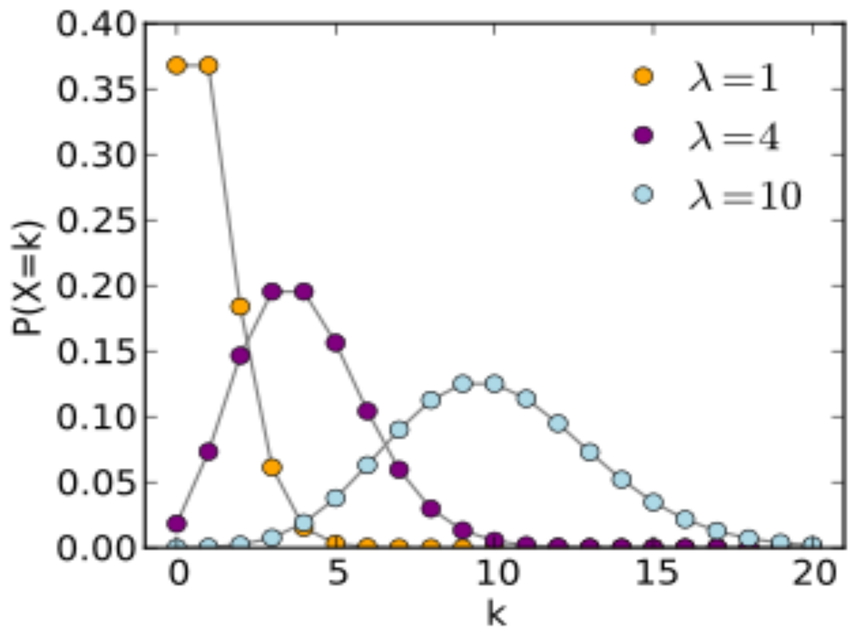
Poisson Approximation

-

Well known: Law of small numbers

n Bernoulli r.v. X_i (p, 1-p), $np = const$

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



Poisson Approximation

- (Somewhat less well-known)

Le Cam's theorem:

n r.v. X_i (with common support $(0,1,2,3,4,\dots)$)

$p_i = \Pr[X_i \neq 0], \lambda = \sum p_i, q_j = \sum \Pr[X_i = j]$

Y_i is a r.v. with distr $(0, \frac{q_1}{\lambda}, \frac{q_2}{\lambda}, \frac{q_3}{\lambda}, \frac{q_4}{\lambda}, \dots \dots)$

Y *is a* ***compound Poisson distr (CPD)***

$$\sum_{i=1}^{N} Y_i \text{ where } N \sim Poisson(\lambda)$$

$$\Delta(\sum X_i, Y) \leq \sum p_i^2$$

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

Poisson Approximation

- **Le Cam's theorem:** $\Delta(\sum X_i, Y) \le \sum p_i^2$
- 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

Outline

- Threshold Probability Maximization
- Stochastic Knapsack
- Other Results

Stochastic Knapsack

- A knapsack of capacity C
- A set of items.
- Known: Prior distr of (size, profit) of each item.
- Items arrive one by one
- Irrevocably decide whether to accept the item
- The actual size of the item becomes known after the decision
- Knapsack constraint: The total size of accepted items <= C
- Goal: maximize E[Profit]

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]

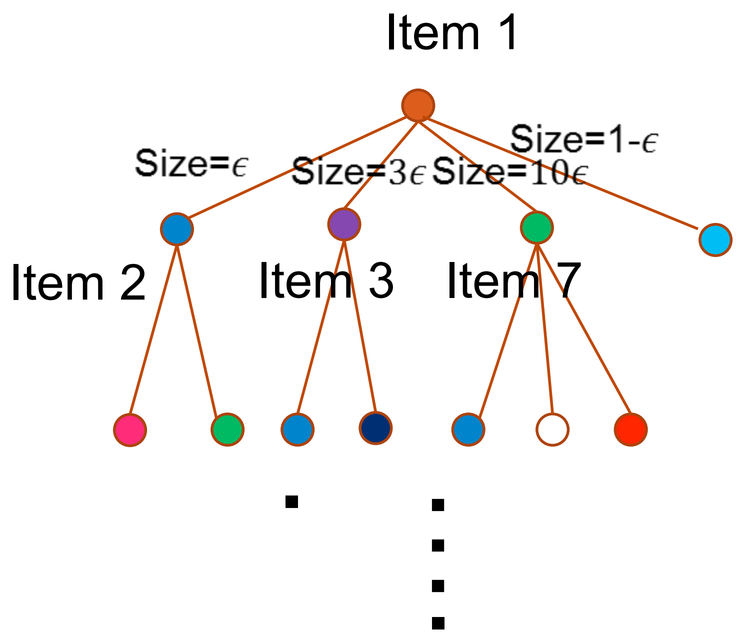
Our result:

$(1+\epsilon, 1+\epsilon)$-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: $(0, \epsilon, 2\epsilon, 3\epsilon, \ldots\ldots, 1)$.
  - All prob. values are in the form of k/M, ($k<M$ and $M=poly(n)$)
  - Profit of each item i is a fixed value

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)

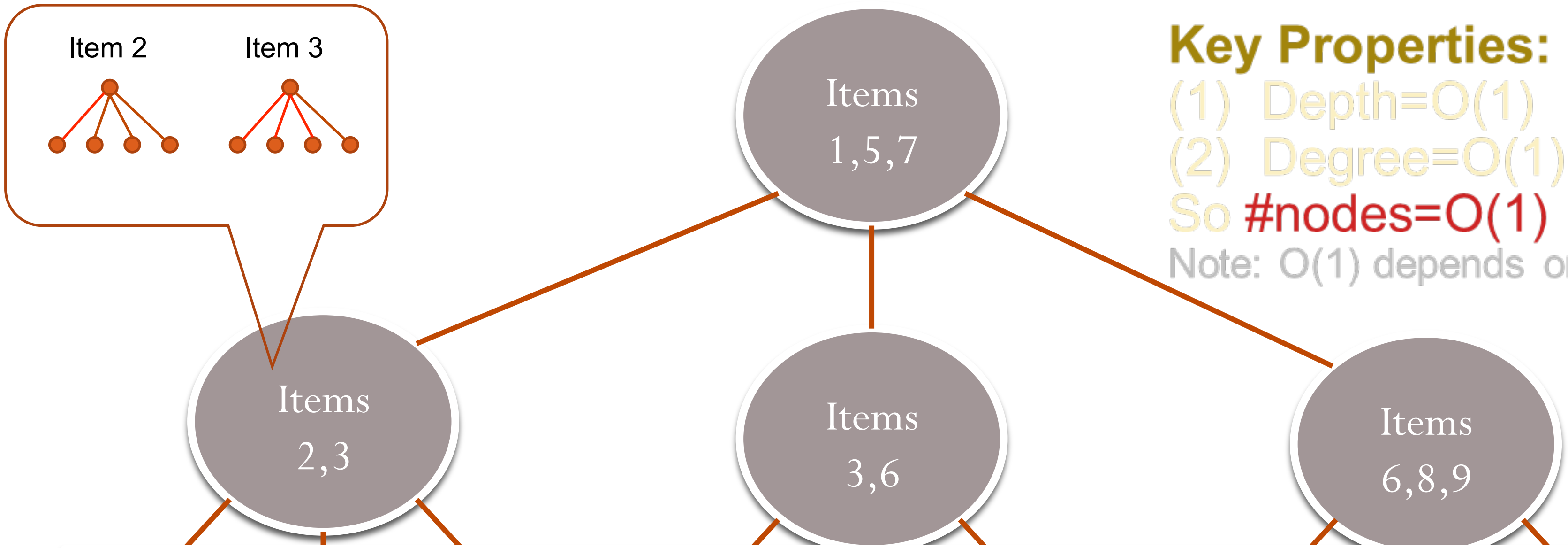
So #nodes=O(1)

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)

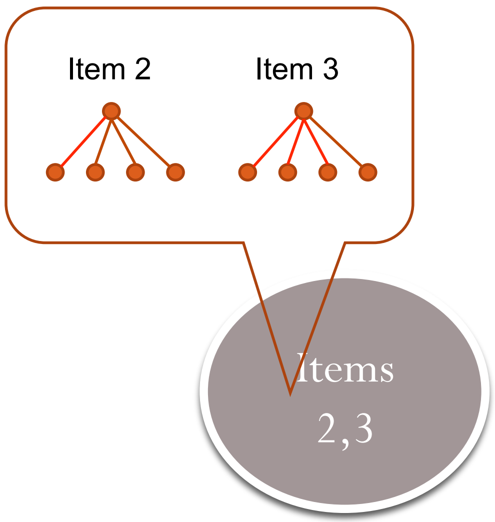
Note: O(1) depends on ϵ

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)

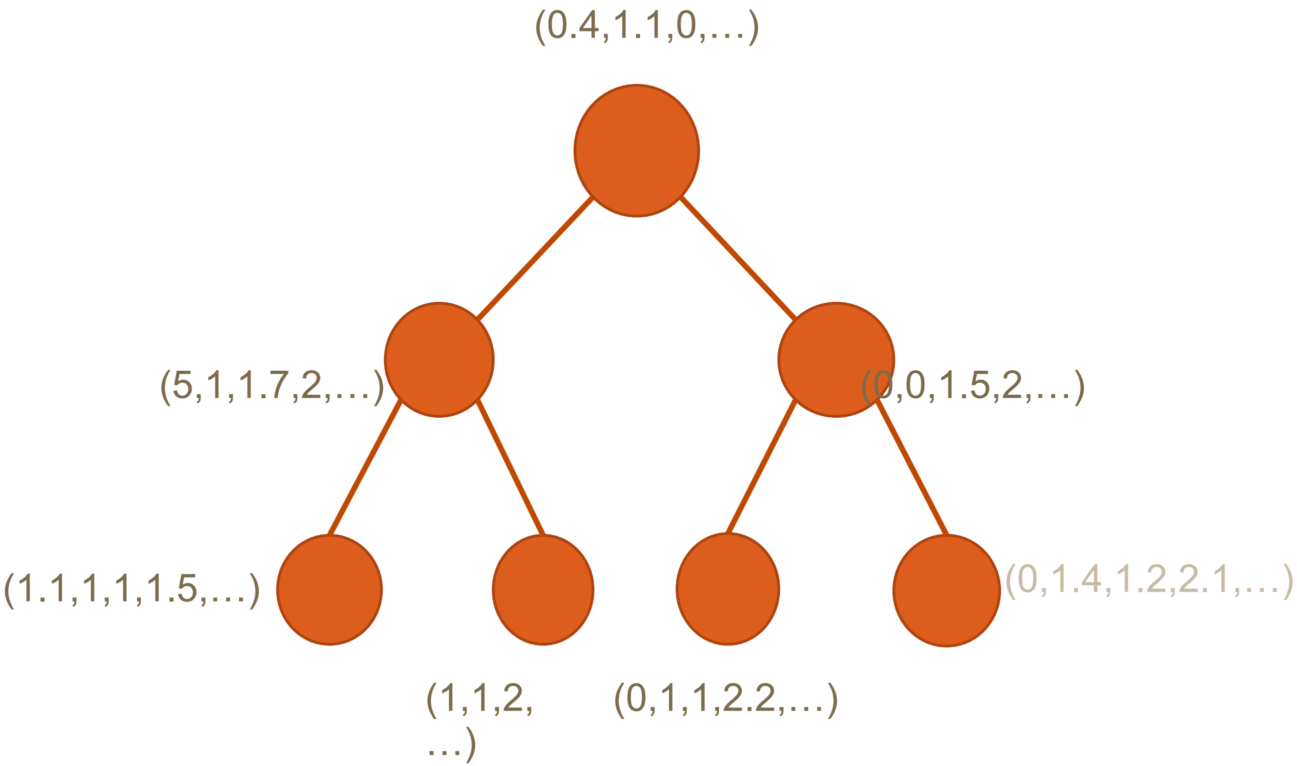


Sg=Sg(*item2*)+Sg(*item3*)

$$\mathrm{CPD}(\mathrm{Sg}) \sim size(item2) + size(item3)$$

Algorithm

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



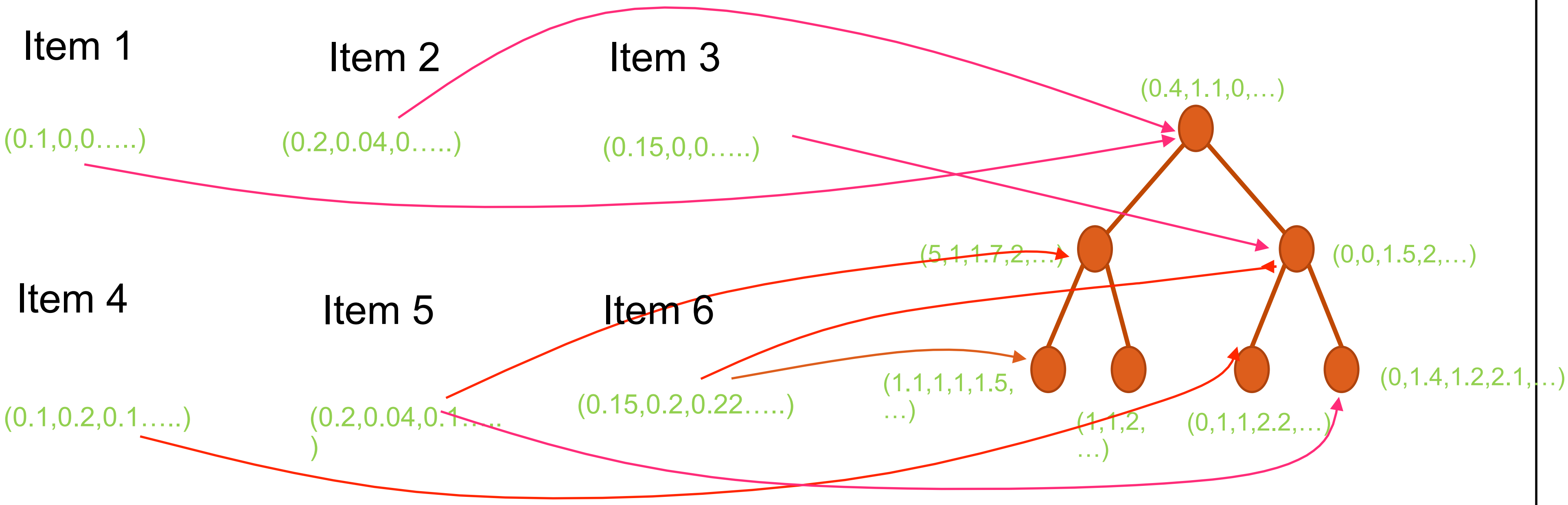
- O(1) nodes
- Poly(n) possible signatures for each node
- So total #configuration =poly(n)

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 program)



On any root-leaf path, we can select one choice for each item

Outline

- Threshold Probability Maximization
- Stochastic Knapsack
- Other Results

Other Results

- 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

Conclusion

- Using Poisson approximation, we can often reduce the stochastic optimization problem to a multi-dimensional packing problem
- More applications

Thanks

lijian83@mail.tsinghua.edu.cn