Two-stage Stochastic Programming

On average:

Minimum Spanning Trees

Abraham Flaxman, CMU

Alan Frieze, CMU

Michael Krivelevich, Tel Aviv

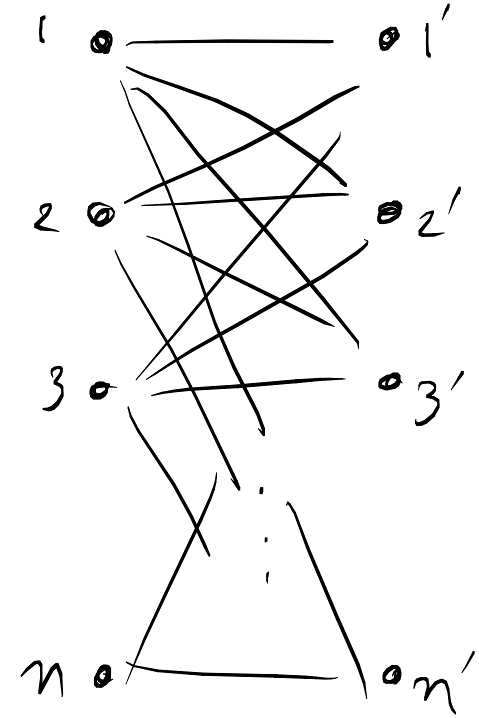
Two-stage Stochastic Programming:

Input: situation today
distribution of situations tomorrow

Output: policy for today & all possible tomorrows.

Objective: Minimize cost today
+ expected cost tomorrow

Example: 2-stage Matching



Input: costs on Monday: c_M

dist. of costs on Tuesday:

$$c_T(e) \sim \mathrm{Exp}(1)$$

Output: Edges to buy on Monday (because policy on Tuesday is obvious — finish the matching.)

Soln: Find opt in poly time by . . . .

Prior Work: (Lots)

http://stoprog.org/

R. Schultz, *Stochastic Integer Programming: A Tutorial*, 9th International Conference on Stochastic Programming, 2001.

M. H. van der Vlerk, *Stochastic Integer Programming Bibliography*, 1996-2003, http://mally.eco.rug.nl/stopbib.html.

Combinatorial Problems:

M. Albareda-Sambola, M. H. van der Vlerk, E. Fernández, Exact solutions to a class of stochastic generalized assignment problem, Stochastic Programming E-Print Series, 2002.

A. Gupta, M. Pal, R. Ravi, A. Sinha, Boosted sampling: Approximation algorithms for stochastic optimization, STOC 2004.

N. Immorlica, D. Karger, M. Minkoff, V. Mirrokni, On the costs and benefits of procrastination: Approximation algorithms for stochastic combinatorial optimization problems, SODA 2004.

N. Kong, A. Schaefer, A factor ½ approximation algorithm for two-stage stochastic matching problems, submitted.

R. Ravi, A. Sinha, Hedging uncertainty: Approximation algorithms for stochastic optimization problems, IPCO 2004.

Not the same model:

H. Ishii, S. Shiode, T. Nishida, Y. Namasuya, Stochastic spanning tree problem, *Discrete Appl. Math.* 3(4):263-273, 1981.

Minimum Spanning Trees:

Input: A graph $G=(V,E)$, a cost vector $c \in \mathbb{R}^E$.

Output: A spanning tree $T \subseteq E$

Objective: $Z_{opt} =$ minimize $\sum_{e \in T} c_e$

Many ways to find optimal value.

$\rightarrow$ Often attributed to Kruskal (1956) and Prim (1957), but dates to Boruvka (1926).

See R. Graham, P. Hell, "History of the min. sp. tree prob.".

Random Minimum Spanning Trees:

If $c_e \sim U[0,1]$ then

as $n \longrightarrow \infty$,

$$z_{opt} \longrightarrow \zeta(3) = \frac{1}{1^3} + \frac{1}{2^3} + \frac{1}{3^3} + \ldots$$

$$\approx 1.20205\ldots$$

[Frieze, 1985]

Sketch of Proof:

length of tree T

$$\ell(T) = \sum_{e \in T} X_e$$

$$= \sum_{e \in T} \int_0^1 \mathbb{1}_{\{X_e \geq p\}} \, dp$$

$$= \int_0^1 \sum_{e \in T} \mathbb{1}_{\{X_e \geq p\}} \, dp$$

for min. sp. tree

$$= \int_0^1 (K(G_p) - 1) \, dp.$$

# of components

$$E[\ell(T)] = \int_{p=0}^1 E[K(G_p)] \, dp - 1$$

only trees, contribute and giants

$$\approx \int_{p=0}^1 \sum \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{kn} \, dp$$

$$\approx \int_{p=0}^1 \sum \frac{n^k}{k!} k^{k-2} p^{k-1} (1-p)^{kn} \, dp$$

$$\approx \sum \frac{n^k}{k!} k^{k-2} \boxed{\int_0^1 p^{k-1} (1-p)^{kn} \, dp}$$

$$\approx \sum \frac{n^k}{k!} k^{k-2} \boxed{\frac{(k-1)!\,(kn)!}{(k(n+1))!}}$$

Beta Integral

$$\approx \sum k^{-3}$$

2-Stage Stochastic Minimum Spanning Trees:

Input: A cost vector c_M

A distribution over cost vectors c_T

Output: A forest F to buy Monday

(since best policy on Tuesday is obvious)

Objective: $$z_{opt} = \text{minimize} \sum_{e \in F} c_M(e) + E\left[\min_{F'} \left\{ \sum_{e \in F'} c_T(e) : F \cup F' \text{ is a spanning tree} \right\}\right]$$

NP-hard, hard to approx. beyond $O(\log n)$.

Random 2-Stage Stochastic Min. Sp. Trees:

$$c_M(e) \sim U[0,1]$$

$$c_T(e) \sim U[0,1]$$

Same question.

Some observations:

① Buying a spanning tree all on Monday costs $\zeta(3)$.

② If you knew Tuesday's costs on Monday, could get $\zeta(3)/2$.

Random 2-Stage Stochastic Min. Sp. Trees:

A reasonable approach:

Pick some threshold value, α.

- On Monday, only buy edges with cost less than α.
- On Tuesday, finish the tree.

Best value $\alpha = \frac{1}{n}$

Objective value $Z_{\text{threshold } \frac{1}{n}} = \zeta(3) - \frac{1}{2}$.

Random 2-Stage Stochastic Min. Sp. Trees:

- Threshold is not optimal:

  By looking at structure instead of just cost, can improve a little.

  $$z_{opt} \leq \zeta(3) - \frac{1}{2} - 10^{-256}.$$

- There is no way to attain $\zeta(3)/2$, because you must make some mistakes on Monday.

  $$z_{opt} \geq \zeta(3)/2 + 10^{-5}.$$

Random 2-Stage Stochastic Min. ~~Sp. Trees~~:

Rooted Arborescence
(directed version of sp. tree)

Threshold heuristic now is:

- On Monday, for all v, if cheapest out-edge costs less than α and does not form a cycle, buy it.
- On Tuesday, complete Rooted Arbor. as cheaply as possible.

Best $\alpha = \frac{1}{n-1}$, which makes $Z_{opt} = 1 - \frac{1}{e}$

Random 2-Stage Stochastic Min. Rooted Arb:

- For rooted Arborescence, threshold heuristic is optimal.

Pf: Every vertex besides the root has to have an out-edge.

Just look at a single vertex.

The expected cost of the min. out-edge on Tuesday is $\frac{1}{n}$.

So the expected cost of the out-edges is at least

$$(n-1)\, E\left[\min\left\{\frac{1}{n}, c_M(e_1), c_M(e_2), \ldots, c_M(e_{n-1})\right\}\right]$$

$$\approx 1 - \frac{1}{e}.$$

One more thing:

Concentration: Say Z = cost of MST on rand. $[0,1]$ edge wts.

Want a bound like

$$\Pr\left[\,\left|Z - \zeta(3)\right| \geq \lambda\,\right] \leq e^{-\delta n}.$$

where (since $E[Z] = \zeta(3)$ is constant) we have λ & δ constant as well.

We prove this using a great inequality:

Symmetrized Logarithmic Sobolev Ineq.:

$Z = f(X_1, \ldots, X_N)$ where all X_i i.i.d.

$Z_i' = f(X_1, \ldots X_i', \ldots, X_N)$ where X_i' is indep copy of X_i.

If $E\left[\sum_{i=1}^{N} (Z - Z_i')^2 \mathbb{1}_{Z > Z_i'} \middle| X_1, \ldots, X_N\right] \leq c$,

then for all $t > 0$,

$$\Pr[Z > E[Z] + t] \leq e^{-t^2/4c}.$$