Quan Nguyen
quan@wustl.edu
March 27, 2020

Agenda:

- Section 1: a slightly tighter bound for PoF(*Leximin*) when not all resources are unique
- Section 2: some results from the Leximin allocation on the real data set
- Section 3: distributions of costs in the real data set
- Section 4: some known result related to the asymptotic behaviors of the random assignment problem

1 Tighter upper-bound for PoF

To illustrate that PoF(*Leximin*) can be arbitrarily close to $n-1$, we use the following cost matrix:

	1	**2**	**3**	**...**	**n - 1**	**n**
1	0	1	1	...	1	1
2	0	1 - ϵ	1	...	1	1
3	0	0	1 - ϵ	...	1	1
...	...	...	...	...	...	...
n - 1	0	0	0	...	1 - ϵ	1
n	0	0	0	...	0	1 - ϵ

Table 1: Cost matrix with PoF = $n-1$

This setup requires that all of the n resources are unique (since no two columns share the same cost vector). This might not be the case in many situations.

Assume we have $m < n$ unique types of resources, each of which has a capacity constraint and these constraints sum to n. Specifically, denote the capacity for resource j as c_j, so resource j can offer at most c_j instances of its service. We also have $\sum_j c_j = n$.

We now prove that under the leximin allocation, a group of any given m agents, each of whom is assigned to a unique resource type, always contains at least one agent with cost 0.

Assume that this is not the case, we see that each of these m agents is not being assigned to their best allocation (since the best allocation for an agent always has cost 0). We thus construct a directed graph with vertices 1, 2, ..., m. This graph has an edge (i, j) when agent

i can get a lower cost than under leximin if assigned to resource j.

Since we assume that no agent is assigned to their best resource, each vertex in this graph has at least one outgoing edge. A directed cycle thus exists, which allows us to implement a "switch" in the current allocation where every agent i is now assigned to resource j, for every edge (i, j) in this directed cycle. After this switch, no agent suffers from a higher cost, while at least one agent will achieve a lower cost. This contradicts the definition of the leximin allocation.

So under leximin, in a group of m agents, each of whom is assigned to a unique resource, there always exists an agent with cost 0. In total, we can construct $\min_j c_j$ such groups that are mutually exclusive and therefore obtain at least $\min_j c_j$ agents with cost 0 under leximin. This will lead to the following inequality:

$$\text{PoF}(\textit{Leximin}) \geq n - \min_j c_j$$

When there is a resource that can only offer one unit of capacity, this bound becomes the original $n - 1$ term.

Note: This bound might be able to be made tighter as there does not seem to be a sample cost matrix where $\min_j c_j > 1$ and equality is achieved.

2 Leximin allocation on the real data set

The previous naive implementation of the leximin allocation optimizes the next leximin value using a binary search. Faster computation can be achieved by adding the leximin constraints and leaving the optimization to the LP solver, Gurobi. (Gurobi's tolerance for objective optimization and constraints is 10^{-4} [1], so the values in the real data set were multiplied by 10^9 and rounded before being put through the solver.)

The leximin allocation on the real data set yields a sum of probability increases of 1169.777 and a sum of probability of 3667.758. This is to be contrasted to the corresponding statistics from the efficient allocation minimizing sum objective: a sum of probability increases of 1129.065 and a sum of probability of 3627.047.

The below figure compares the distribution of costs (probability increases) under the two allocations:

[1] https://www.gurobi.com/documentation/9.0/refman/mipgap2.htmlparameter:MIPGap

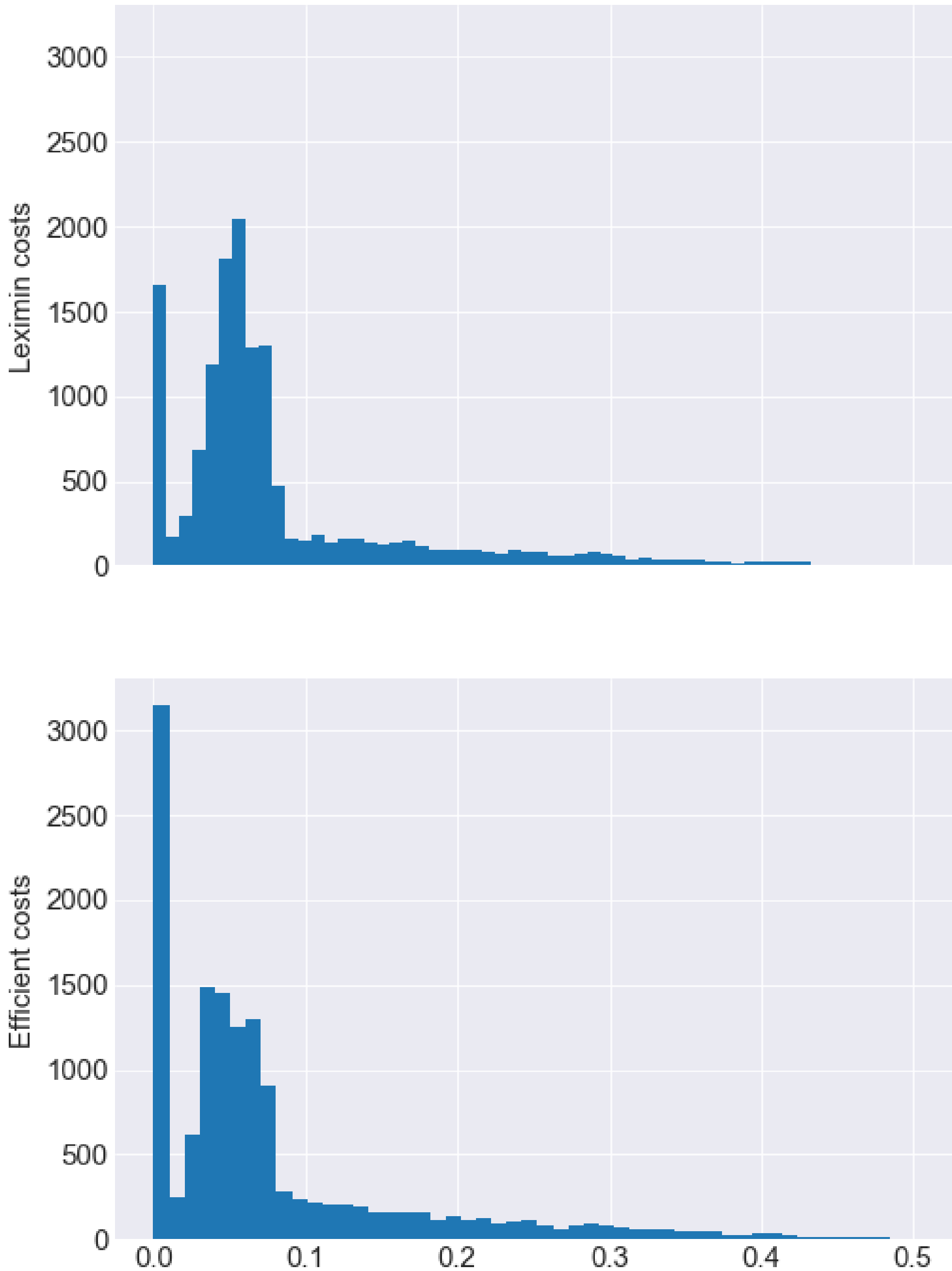

Figure 1: Assigned probability increase distribution

3 Probability increases in the real data set

The following histogram denotes the distribution of the original probability increases in the real data set:

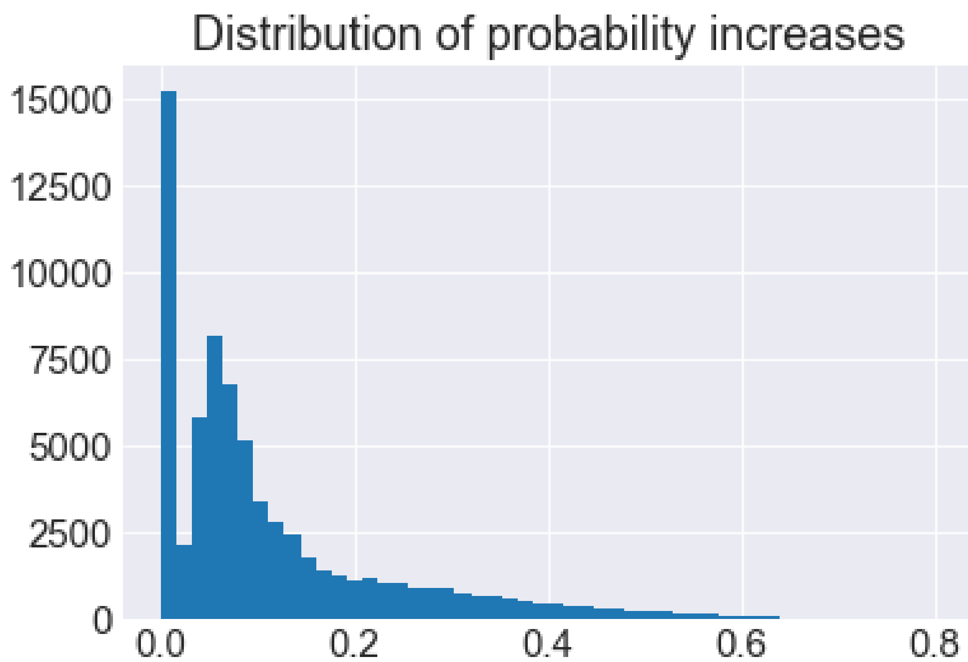

Figure 2: Probability increase distribution

However, individual values in our data might not be i.i.d. with respect to a common probability distribution. The values in individual columns, representing how good individual intervention are, have distributions different from the overall distribution above:

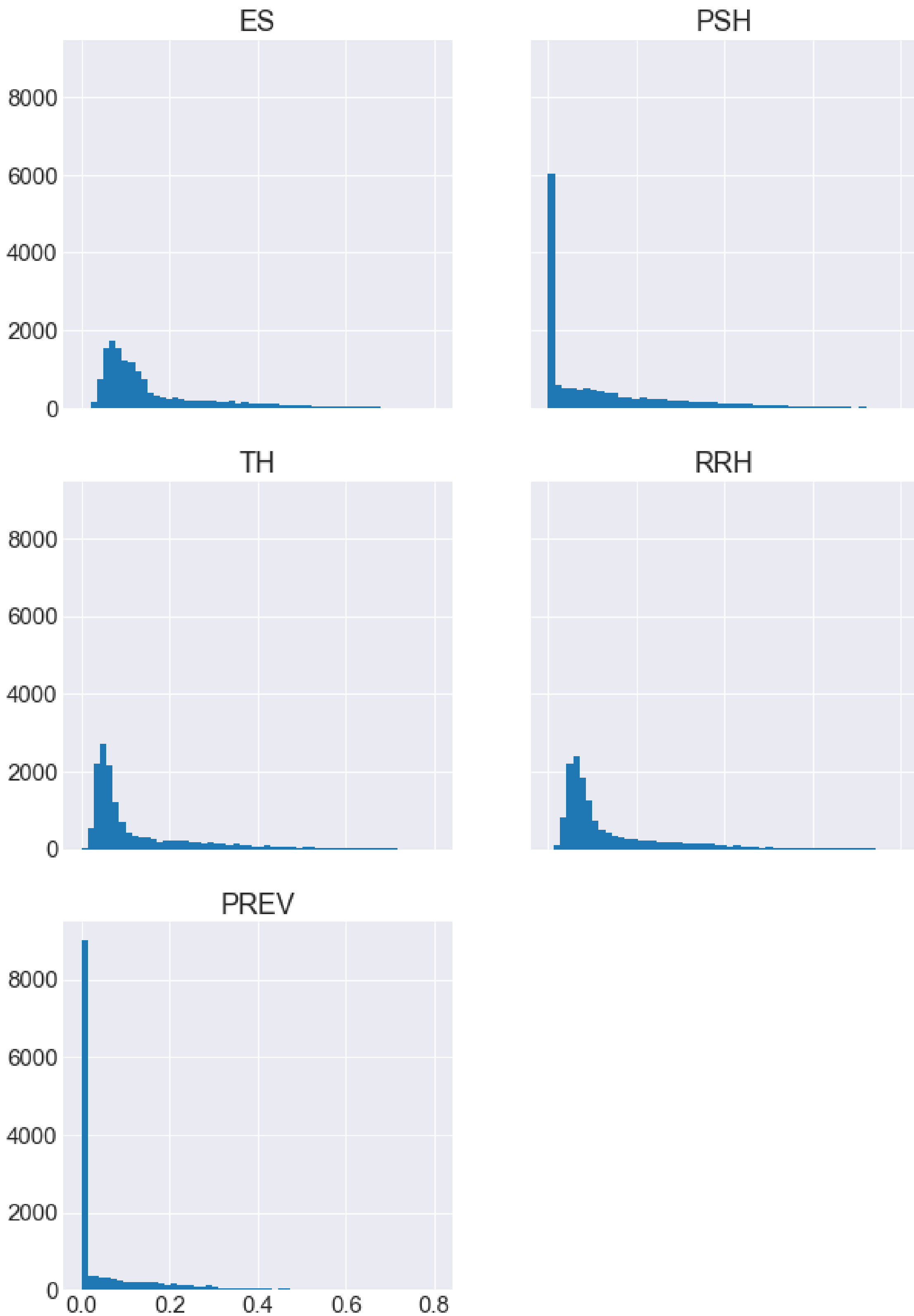

Figure 3: Cost distributions of different intervention

Values in one column might also be correlated with those in another column. The following is the correlation matrix of the costs (probability increases) among the five interventions:

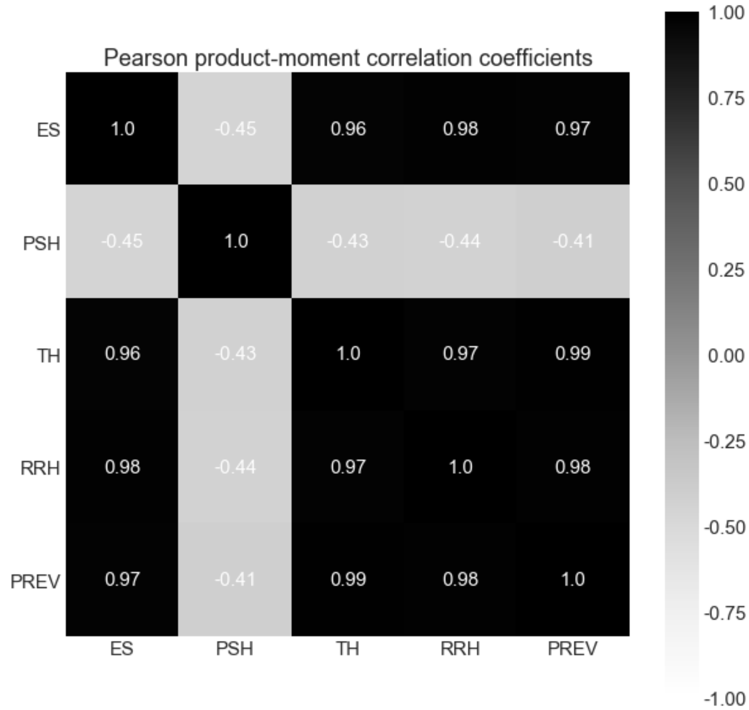

Figure 4: Correlation matrix of costs

Overall, the individual costs from our real data might not be i.d.d. from a common probability distribution.

4 Asymptotic behaviors of the random assignment problem

4.1 The efficient allocation

We denote $\mathbb{E}[c^*] = \lim_{n\to\infty} \mathbb{E}[c_n]$, where c_n is the optimal total cost of a given instance of the assignment problem given by an $n \times n$ cost matrix.

Frenk et al. [1987] showed that if c_{ij} are drawn from a distribution function F, which is defined on $(0, \infty)$ and satisfies $\lim_{n\to\infty} F^{-1}(1/n) = 0$, then:

$$0 \leq \liminf_{n\to\infty} \frac{\mathbb{E}[c_n]}{n\ F^{-1}(1/n)} \leq \limsup_{n\to\infty} \frac{\mathbb{E}[c_n]}{n\ F^{-1}(1/n)} < \infty$$

Olin [1992] used some stronger assumptions on F to derive the following bound that holds almost surely:

$$\frac{1+e^{-1}}{f(0^+)} \leq \liminf_{n\to\infty} c_n \leq \limsup_{n\to\infty} c_n \leq \frac{4}{f(0^+)}$$

where f is the continuous density function of F is that positive in the neighborhood of 0.

Perhaps most famously, when the costs are i.d.d. from the uniform distribution with support for $[0,1]$ or the exponential distribution with mean 1, we have the following analytical expression:

$$\mathbb{E}[c_n] = \sum_{i=1}^{n} \frac{1}{i^2}$$

This equation was independently proven by Linusson and Wästlund [2004] and Nair et al. [2005]. Taking the limit when $n \to \infty$, we have:

$$\mathbb{E}[c^*] = \sum_{i=1}^{\infty} \frac{1}{i^2} = \zeta(2) = \frac{\pi^2}{6}$$

Additionally, Aldous [2001], who also proved the above $\zeta(2)$ limit, noted that the result holds for distributions with the same density at 0 (i.e., $f(0) = 1$).

4.2 The bottleneck allocation

The literature on the asymptotic behavior of the leximin allocation mostly consists of the analysis of $\mathbb{E}(b_n)$, where b_n is the optimal largest cost (or bottleneck) resulting from a bottleneck assignment of an $n \times n$ cost matrix.

Pferschy [1996] showed that the expected optimal bottleneck approaches the left endpoint of the support of the distribution F of the costs:

$$\lim_{n\to\infty} \mathbb{E}[b_n] = \inf\{x|F(x) > 0\} \tag{1}$$

if the support of F is upper-bounded: $\sup\{x|F(x) < 1\} < +\infty$.

Additionally, if the costs are i.d.d. random variables drawn from $U[0,1]$, Pferschy [1996] also showed that:

$$\mathbb{E}[b_n] < 1 - \left(\frac{2}{n(n+2)}\right)^{2/n} \frac{n}{n+2} + \frac{123}{610n}, \forall n > 78,$$

$$\mathbb{E}[b_n] \leq \frac{\ln(n) + c}{n} + O\left(\frac{\ln^2(n)}{n^2}\right), \text{where } c \text{ is a constant}$$

Spivey [2011] considered asymptotic moments of b_n. It was shown that even for certain distributions that are not upper-bounded such as the exponential, the half-normal, and the Pareto, the result in Equation (1) still holds. More importantly, assuming $Q = F^{-1}$, $Q(0) = 0$, and Q can be expanded in a Maclaurin series, this work derives the following expression for $p \geq 1$:

$$E\left[(b_n^*)^p\right]$$
$$= \begin{cases} Q'(0)\left(\frac{\log n}{n} + \frac{\log 2+\gamma}{n}\right) + o\left(\frac{(\log n)^2}{n^{7/5}}\right), & mp = 1; \\ \left(\frac{Q^{(m)}(0)}{m!}\right)^p \left(\frac{(\log n)^2}{n^2} + \frac{2(\log 2+\gamma)\log n}{n^2} + \frac{\zeta(2)+\gamma^2+2\gamma\log 2-(\log 2)^2}{n^2}\right) + o\left(\frac{(\log n)^2}{n^{7/3}}\right), & mp = 2; \\ \left(\frac{Q^{(m)}(0)}{m!}\right)^p \left(\frac{(\log n)^{mp}}{n^{mp}} + \frac{mp(\log 2+\gamma)(\log n)^{mp-1}}{n^{mp}}\right) + o\left(\frac{(\log n)^{mp-2}}{n^{mp}}\right), & mp \geq 3. \end{cases}$$

where $m = \min_{d\geq 0}\{Q^{(d)}(0) \neq 0\}$.

In the specific cases of the $U[0,1]$ distribution or the exponential distribution with mean 1:

$$\mathbb{E}[b_n] = \frac{\log(n) + \log(2) + \gamma}{n} + O\left(\frac{\log^2(n)}{n^{7/5}}\right)$$

References

Johannes Bartholomeus Gerardus Frenk, M Van Houweninge, and AHG Rinnooy Kan. Order statistics and the linear assignment problem. *Computing*, 39(2):165–174, 1987.

Birgitta Olin. *Asymptotic properties of random assignment problems*. Royal Inst. of Technology, Department of Mathematics, Division of Optimization and Systems Theory, 1992.

Svante Linusson and Johan Wästlund. A proof of parisi's conjecture on the random assignment problem. *Probability theory and related fields*, 128(3):419–440, 2004.

Chandra Nair, Balaji Prabhakar, and Mayank Sharma. Proofs of the parisi and coppersmith-sorkin random assignment conjectures. *Random Structures & Algorithms*, 27(4):413–444, 2005.

David J Aldous. The $\zeta(2)$ limit in the random assignment problem. *Random Structures & Algorithms*, 18(4):381–418, 2001.

Ulrich Pferschy. The random linear bottleneck assignment problem. *RAIRO-Operations Research*, 30(2):127–142, 1996.

Michael Z Spivey. Asymptotic moments of the bottleneck assignment problem. *Mathematics of Operations Research*, 36(2):205–226, 2011.