

# The College Application Problem

Max Kapur

Seoul National University  
Department of Industrial Engineering  
Management Science/Optimization Lab

May 3, 2022

# Introduction

The optimal college application problem is a **novel combinatorial optimization problem**.

It involves **maximizing the expected maximum** of a portfolio of random variables subject to a budget constraint.

Methodological orientation:

- Investment with uncertain payoff, search for the efficient frontier recall **portfolio allocation** models.
- Generalizes the **knapsack problem**: Integral packing constraint, NP-completeness, approximation algorithms.
- Objective is a **submodular set function**.

Today's presentation: **define the problem** and briefly summarize our **solution algorithms**.

# The admissions process

Consider a **single student's** college application decision.

Market contains  $m$  **schools**, indexed by  $\mathcal{C} = \{1 \dots m\}$ . School  $j$  is named  $c_j$ .

Given a student's academic records, test scores, and demographic information, we can estimate her **admissions probability**  $f_j$  at each school.

Let the **random variable**  $Z_j \sim \text{Bernoulli}(f_j) = 1$  if student is admitted, 0 otherwise. Assume independent.

Let  $\mathcal{X} \subseteq \mathcal{C}$  denote the set of schools, or **application portfolio**, to which a student applies.

Application fees, time to write essays, and/or legal limits **constrain** applicant behavior. We consider a single knapsack constraint  $\sum_{j \in \mathcal{X}} g_j \leq H$  where  $g_j$  is called  $c_j$ 's **application cost**.

# Utility model

Let  $t_j \geq 0$  denote the **utility** the student receives if she attends  $c_j$ . Wlog,  $t_j \leq t_{j+1}$ .

Let  $t_0$  denote her utility if she doesn't get into college. Wlog,  $t_0 = 0$  (see paper).

The student's overall utility is the  $t_j$ -value associated with the **best** school she gets into:

$$\text{Utility} = \max\{t_0, \max\{t_j Z_j : j \in \mathcal{X}\}\}$$

When the student's application portfolio is  $\mathcal{X}$ , we refer to her expected utility as the portfolio's **valuation**  $v(\mathcal{X})$ .

# Unpacking the portfolio valuation function

To get  $v(\mathcal{X})$  into a tractable form, let  $p_j(\mathcal{X})$  denote the probability that the student **attends**  $c_j$ .

This happens if and only if she **applies** to  $c_j$ , is **admitted** to  $c_j$ , and is **not admitted** to any school she prefers to  $c_j$ :

$$p_j(\mathcal{X}) = \begin{cases} f_j \prod_{\substack{i \in \mathcal{X}: \\ i > j}} (1 - f_i), & j \in \{0\} \cup \mathcal{X} \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$v(\mathcal{X}) = \sum_{j=1}^m t_j p_j(\mathcal{X}) = \sum_{j \in \mathcal{X}} \left( f_j t_j \prod_{\substack{i \in \mathcal{X}: \\ i > j}} (1 - f_i) \right).$$

# Problem statement

## Problem 1 (The college application problem)

$$\begin{aligned} \text{maximize} \quad & v(\mathcal{X}) = \sum_{j \in \mathcal{X}} \left( f_j t_j \prod_{\substack{i \in \mathcal{X}: \\ i > j}} (1 - f_i) \right) \\ \text{subject to} \quad & \mathcal{X} \subseteq \mathcal{C}, \quad \sum_{j \in \mathcal{X}} g_j \leq H \end{aligned}$$

## Problem 2 (The college application problem, INLP form)

$$\begin{aligned} \text{maximize} \quad & v(x) = \sum_{j=1}^m \left( f_j t_j x_j \prod_{i>j} (1 - f_i x_i) \right) \\ \text{subject to} \quad & x_j \in \{0, 1\}, j \in \mathcal{C}; \quad \sum_{j=1}^m g_j x_j \leq H \end{aligned}$$

## Safety, Target, & Reach Schools: How to Find the Right Ones

### What's Covered:

- [What Are Reach, Target, and Safety Schools?](#)
- [Factors that Impact Your Chances](#)
- [Elements of a Balanced College List](#)

Creating a school list is an important-yet-tricky step in the college application process. A strategically constructed school list weighs your desire to attend reach schools—the institutions you dream about going to—along with safety schools where you're very likely to secure admission. Consequently, the ideal school list is balanced between reach, target, and safety schools, allowing you to shoot for the stars while also ensuring admission into at least one school.

### What Are Reach, Target, and Safety Schools?

"Reach," "safety," and "target" are common terms used in college applications to describe the odds a student has of getting accepted at a particular institution. Understanding these terms, and which categories colleges fall into, is a critical step in the application process.

### What is a Reach School?

Reach schools are colleges where you have less than a 25% chance of admission [this is your estimated chance of non-acceptance, not the probability of acceptance and, thus, is much more complex].

## [대입 수시 전략] 총 6번의 기회 ... '상향·소신·안정' 분산 지원하라

중앙일보 | 업데이트 2015.08.26 10:15

저녁보기 ①

전면회 기자

구독

대학 최저학력기준 고려해 전략 지원  
지난해 같은 전형 합격한 선배 내신 참고  
수능 전 대학별고사 보는 곳은 최소화

'지피지기 백전불태(知彼知己百戰不殆)'. 적을 알고 나를 알면 백 번 싸워도 위태롭지가 않다는 뜻이다. 고대 중국의 병법서인 『손자』에 나온 말이지만 현대사회에서도 여러 가지 분야에서 회자된다. 그중 하나가 대학입시다. 특히 2주 앞으로 다가온 수시모집은 전형 종류가 다양해 '적'(모집전형)을 알고, '나'(학생)에 대해 파악하는 게 무엇보다 중요하다.

자신의 학교생활기록부, 교과성적, 대학별고사 준비 상황, 예상 수능점수, 최저학력기준 통과 가능성에 대해 자세히 살핀 후 지원해야 합격률을 높일 수 있다. 수시모집 마무리 전략을 알아봤다.

논술전형을 학생부 성적 기준으로 지원

Do you trust the admissions consultant's advice?

# Existing solutions

In practice, most use **distributive heuristics** such as distributing applications evenly among reach, target, and safety schools. Turns out to be a **risk-averse** approach.

Another intuitive idea is the **linearization heuristic**. Since the expected utility associated with applying to  $c_j$  (alone) is  $f_j t_j$ , solve the knapsack problem

$$\text{maximize } \sum_{j \in \mathcal{X}} f_j t_j \quad \text{subject to } \sum_{j \in \mathcal{X}} g_j \leq H$$

as a surrogate. But this solution can be **arbitrarily bad**.

Fu (2014) solved a similar problem by **enumeration**, which is intractable for  $m \geq 20$  or so.

Our algorithms provide both **time and accuracy guarantees**.



# Our algorithms

In the **special case** where each  $g_j = 1$ , we provide an  $O(m^2)$  algorithm.

The general problem is **NP-complete** (reduction from knapsack). We offer four algorithms:

- A linear relaxation and **branch-and-bound scheme**. Primarily of theoretical interest.
- A **dynamic program based on total expenditures**. Exact solution in  $O(Hm + m \log m)$  time (pseudopolynomial). Very fast for “typical” instances in which  $g_j$  are small integers.
- A different DP based on truncated portfolio valuations.  $(1 - \varepsilon)$ -optimal solution in  $O(m^3/\varepsilon)$  time: an **FPTAS**!
- A **simulated annealing heuristic**. Fast, typically within 2% of optimality.

# Conclusion

“Maximax” form, integrality constraints make the college application problem **theoretically interesting**. Formally, it is a submodular maximization problem, but its approximability is more like knapsack.

Solutions to the college application problem have **practical value**: US admissions consultants charge an average of \$200/hr!  $\Rightarrow$  open-sourcing our code is in the public interest.

Lots of extensions to consider: parametric risk aversion, distribution constraints, memory-usage improvements.

# References

- Fu, Chao. 2014. “Equilibrium Tuition, Applications, Admissions, and Enrollment in the College Market.” *Journal of Political Economy* 122 (2): 225–81. <https://doi.org/10.1086/675503>.
- Kapur, Max. 2022. “OptimalApplication.” GitHub repository. <https://github.com/maxkapur/OptimalApplication>.
- Kulik, Ariel, Hadas Shachnai, and Tami Tamir. 2013. “Approximations for Monotone and Nonmonotone Submodular Maximization with Knapsack Constraints.” *Mathematics of Operations Research* 38 (4): 729–39. <https://doi.org/10.1287/moor.2013.0592>.
- Sklarow, Mark. 2018. *State of the Profession 2018: The 10 Trends Reshaping Independent Educational Consulting*. Technical report, Independent Educational Consultants Association. <https://www.iecaonline.com/wp-content/uploads/2020/02/IECA-Current-Trends-2018.pdf>.