The College Application Problem

Max Kapur and Sung-Pil Hong

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

May 2, 2022

Introduction

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

Maximize the expected maximum of a portfolio of random variables subject to a budget constraint.

Today's agenda:

- **Formulate** college application as an optimization problem.
- The **status quo** in admissions consulting, and why common heuristics fall short.
- Our theoretical results, **solution algorithms**, and Julia implementation.

Proofs and additional results in our arXiv paper (Kapur and Hong 2022).

Methodological orientation

College application spans several areas of combinatorial optimization:

- Investment with uncertain payoff, search for the efficient frontier recall **portfolio allocation** models.
- Generalizes the **knapsack problem**: Integral packing constraint, NP-completeness.
- Objective is a **submodular set function**, but approximation results suggest college application is a relatively easy instance.

Model

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** $Z_j \sim \mathrm{Bernoulli}(f_j) = 1$ 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:

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

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,

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

Problem statement

Problem 1 (The college application problem)

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

Problem 2 (The college application problem, INLP form)

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

The status quo

Timothy Peck — May 7, 2021 — 5 — 12th Grade, College Lists

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

The JoongAng

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

중앙일보 | 업데이트 2015.08.26 10:15 지면보기

전민희 기자 구독

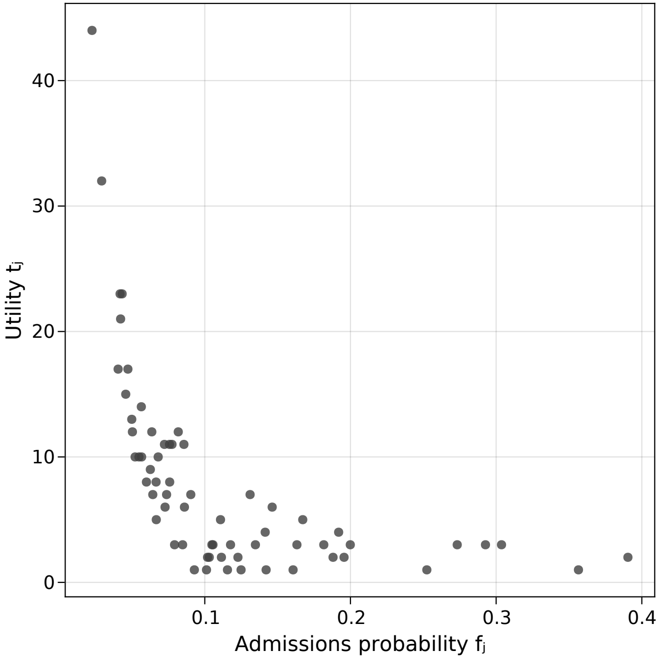
대학 최저학력기준 고려해 전략 지원
지난해 같은 전형 합격한 선배 내신 참고
수능 전 대학별고사 보는 곳은 최소화

'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.

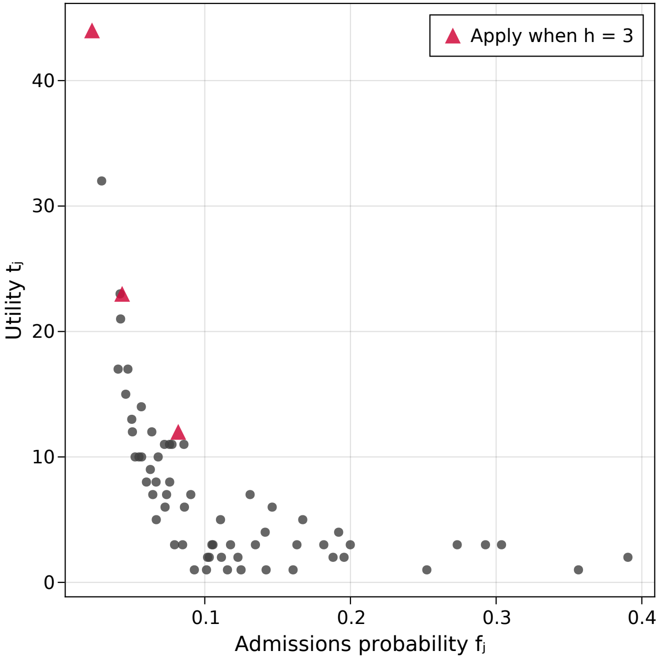
자신의 학교생활기록부, 교과성적, 대학별고사 준비 상황, 예상 수능점수, 최저학력기준 통과 가능성에 대해 자세히 살핀 후 지원해야 합격률을 높일 수 있다. 수시모집 마무리 전략을 알아봤다.

논술전형도 학생부 성적 기준으로 지원

Can we trust the admissions consultant's advice?



Utility t_j
0
10
20
30
40
0.1
0.2
0.3
0.4
Admissions probability f_j



Apply when h = 3
40
30
20
10
0
Utility t_j
0.1
0.2
0.3
0.4
Admissions probability f_j

Existing solutions

In practice, most use **distributive heuristics** such as distributing applications evenly among attractive, selective "reach schools" and less-attractive "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

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

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

Homogeneous costs: A polynomially solvable case

We first consider the **special case** in which each $g_j = 1$ and H is simply $h \leq m$, a limit on the number of schools you can apply to.

This case mirrors the main Korean admissions cycle, in which $h = 3, m = 202$.

We show that when $\mathcal{X}_h$ is the optimal portfolio at h, $\mathcal{X}_h \subset \mathcal{X}_{h+1}$. This **nestedness property** implies the optimality of a **greedy algorithm** that adds schools one at a time, maximizing $v(\mathcal{X})$ at each addition.

Using a variable-elimination technique, we reduce the cost of computing $v(\mathcal{X})$ to amortized unit time, obtaining an $O(hm)$ algorithm. Our Julia implementation (Kapur 2022) solves an $m = 16384$ instance in 200 ms.

Strengthens the result of Fisher et al. (1978).

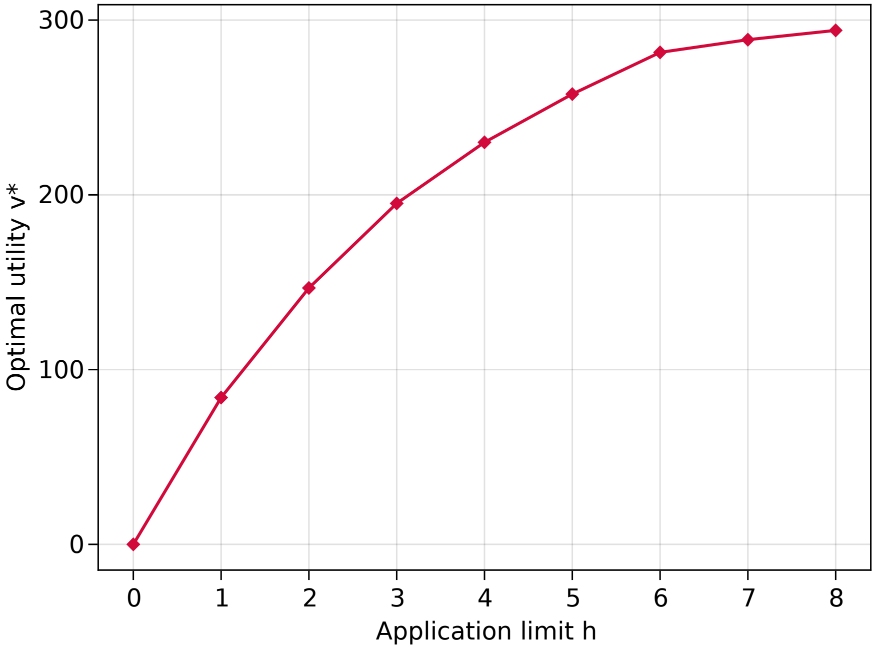
A small example

College data and optimal application portfolios for a fictional market with $m = 8$ schools.

j	**School** c_j	**Admit prob.** f_j	**Utility** t_j	**Priority**	$v(\mathcal{X}_h)$
1	Mercury University	0.39	200	4	230.0
2	Venus University	0.33	250	2	146.7
3	Mars University	0.24	300	6	281.5
4	Jupiter University	0.24	350	1	84.0
5	Saturn University	0.05	400	7	288.8
6	Uranus University	0.03	450	8	294.1
7	Neptune University	0.10	500	5	257.7
8	Pluto College	0.12	550	3	195.1

By the nestedness property, the optimal portfolio when the application limit is h consists of the h schools having priority h or less.

Valuation function appears on next slide. Always concave.



300
200
100
0
Optimal utility v*
0
1
2
3
4
5
6
7
8
Application limit h

Algorithms for the general problem

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.

Existence of FPTAS suggests college application is **a relatively easy instance of submodular maximization** (cf. Kulik et al. 2013).

Conclusion

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.

The nestedness result for the $g_j = 1$ special case also resembles the knapsack problem, although the proof is more subtle.

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 I

Fisher, Marshall, George Nemhauser, and Laurence Wolsey. 1978. "An analysis of approximations for maximizing submodular set functions—I." *Mathematical Programming* 14: 265–94.

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.

Kapur, Max and Sung-Pil Hong. 2022. "The Optimal College Application Problem." Preprint. https://github.com/maxkapur/CollegeApplication.

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.

Appendix: Summary of algorithms

Algorithm	Problem	Restrictions	Exactness	Computation time
Naïve	Homogeneous costs	None	$(1/h)$-opt.	$O(m)$
Greedy	Homogeneous costs	None	Exact	$O(hm)$
Branch and bound	General	None	Exact	$O(2^m)$
Costs DP	General	g_j integer	Exact	$O(Hm + m \log m)$
FPTAS	General	None	$(1-\varepsilon)$-opt.	$O(m^3/\varepsilon)$
Simulated annealing	General	None	0-opt.	$O(Nm)$