Wisdom and Foolishness of Noisy Matching Markets

Kenny Peng and Nikhil Garg
Cornell Tech

arXiv:2402.16771

- In two-sided matching markets, applicants are evaluated noisily
- Do the “correct” matches form?

Our setup

- Colleges share true preferences over students, but rank students noisily
- Assume limited total capacity. Which students match?

Model

- Continuum of students with **true values** v dist. according to η
- College $c \in C$ ranks student using **estimated value** $v + X_c$ $\;(X_c \sim \mathscr{D})$
- Students rank colleges arbitrarily
- Colleges have arbitrary (but not too big) capacities, summing to $S < 1$

In stable matching, what is probability student with true value v matches?

Stable Matching

"Cutoff characterization" of stable matching (Azevedo and Leshno, 2016):

- Student v can **afford** college c iff est. value $v + X_c$ exceeds **cutoff** P_c
- Student matches to favorite college they can afford
- Cutoffs $P_1, P_2, \cdots, P_C$ **market clearing** iff college capacities properly filled
- Lemma: For market-clearing cutoffs, corresponding matching is stable

Key fact

For $P_1, P_2, \cdots, P_C$ market clearing, a student with true value v matches iff $v + X_c > P_c$ for some $c \in C$.

(i.e., the student can afford at least one college.)

Definition 1

A noise distribution $\mathscr{D}$ is **max-concentrating** iff

$$\lim_{n\to\infty} \text{Var}\left[\max\{X_1, X_2, \cdots, X_n\}\right] = 0,$$

for $X_1, X_2, \cdots, X_n \sim \mathscr{D}$.

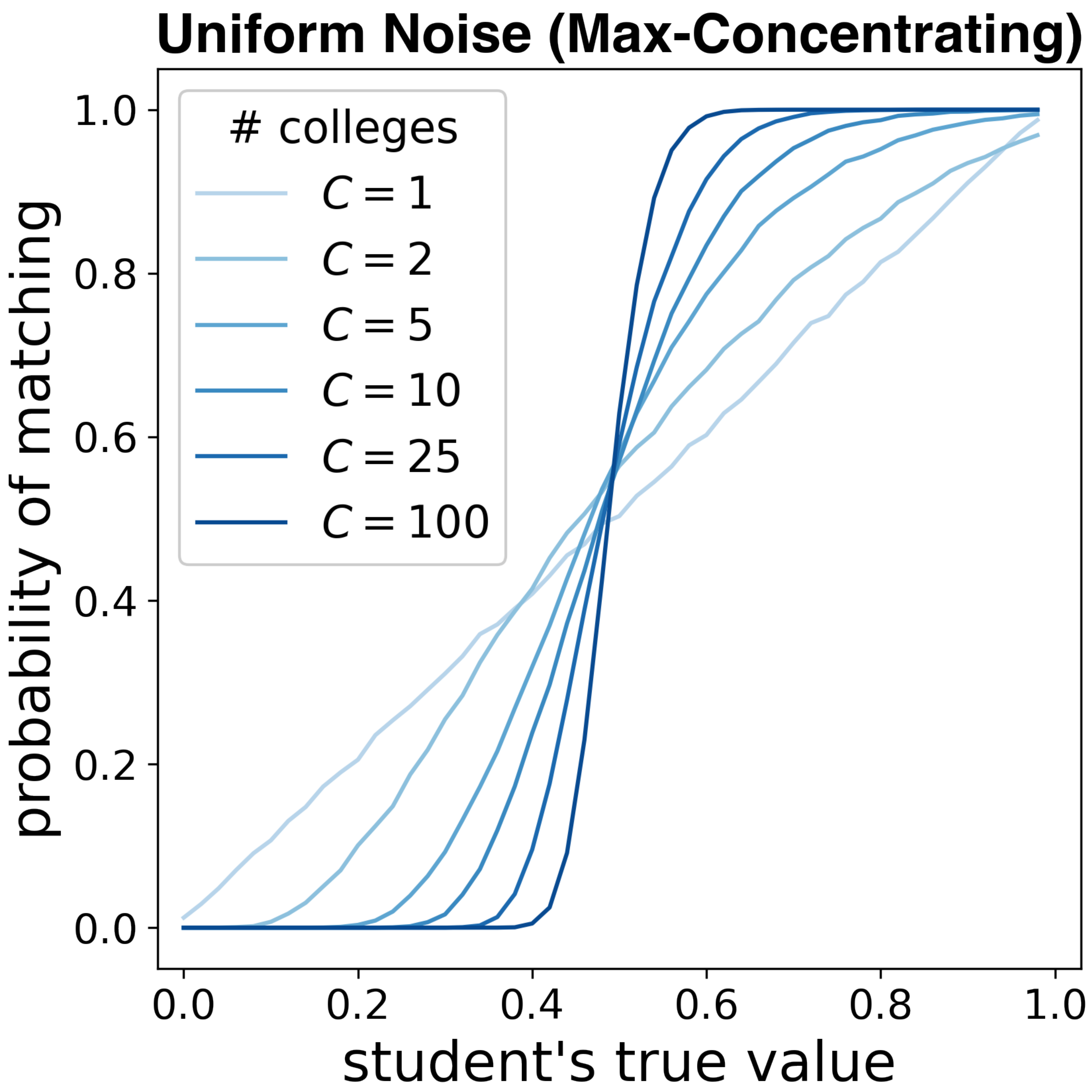
$\mathscr{D}$ is **max-concentrating** iff

$$\lim_{n\to\infty} \text{Var}\left[\max\{\underbrace{X_1, X_2, \cdots, X_n}_{\sim \mathscr{D}}\}\right] = 0$$

Theorem 1 (Informal)

If noise distribution $\mathscr{D}$ is max-concentrating:

$$\lim_{C\to\infty} \Pr\left[\text{student with true value } v \text{ is matched}\right] = \begin{cases} 0 & v < v^* \\ 1 & v > v^* \end{cases}.$$


Uniform Noise (Max-Concentrating)
colleges
$C = 1$
$C = 2$
$C = 5$
$C = 10$
$C = 25$
$C = 100$
probability of matching
student's true value
0.0
0.2
0.4
0.6
0.8
1.0

Definition 2

A noise distribution $\mathscr{D}$ is **long-tailed** iff for all $d > 0$,

$$\lim_{a \to \infty} \Pr_{X \sim \mathscr{D}} [X > a + d \quad X > a] = 1.$$

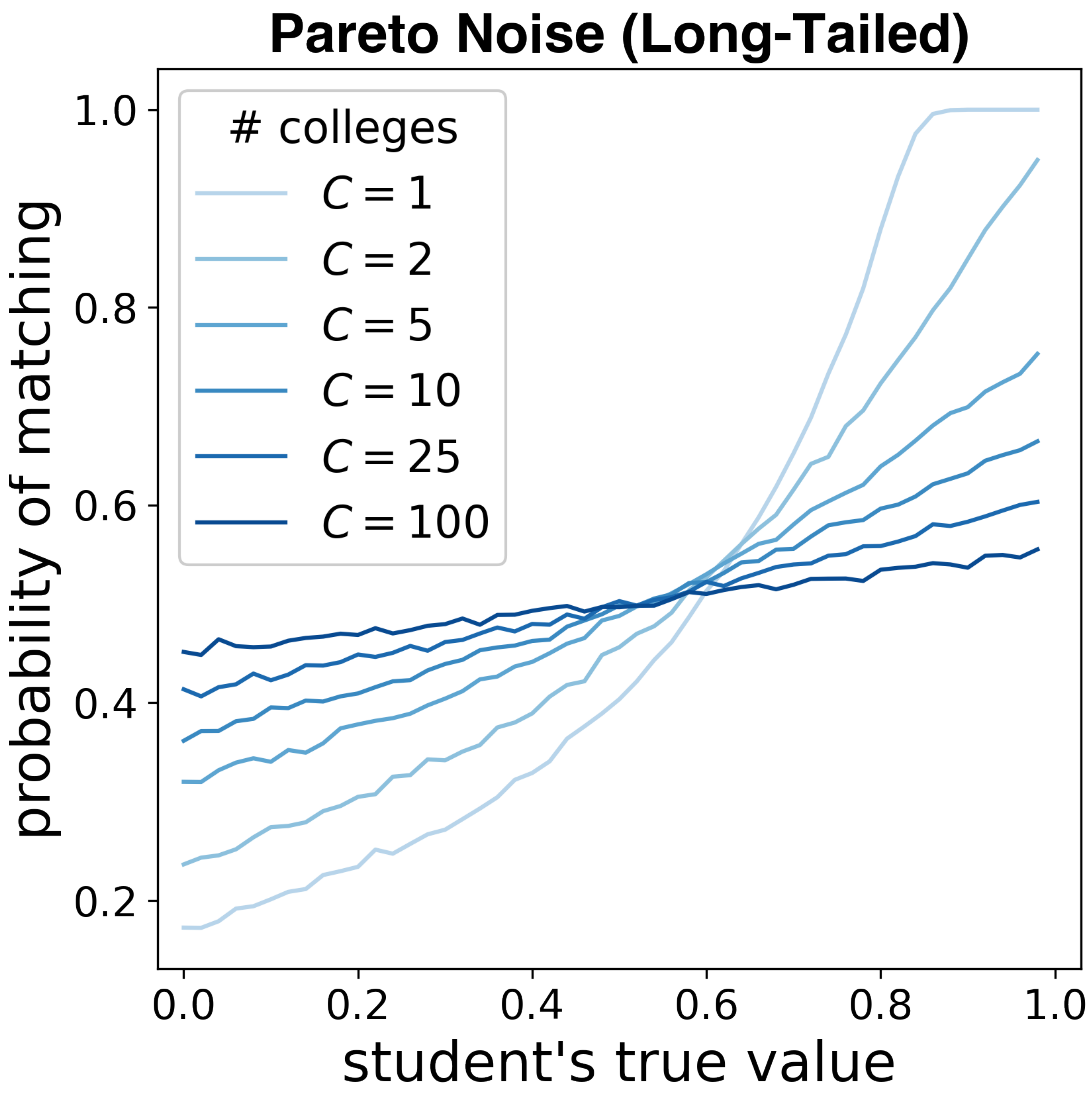
$\mathscr{D}$ is **long-tailed** iff for all $d > 0$,

$$\lim_{a \to \infty} \Pr_{X \sim \mathscr{D}} \left[X > a + d \quad X > a \right] = 1$$

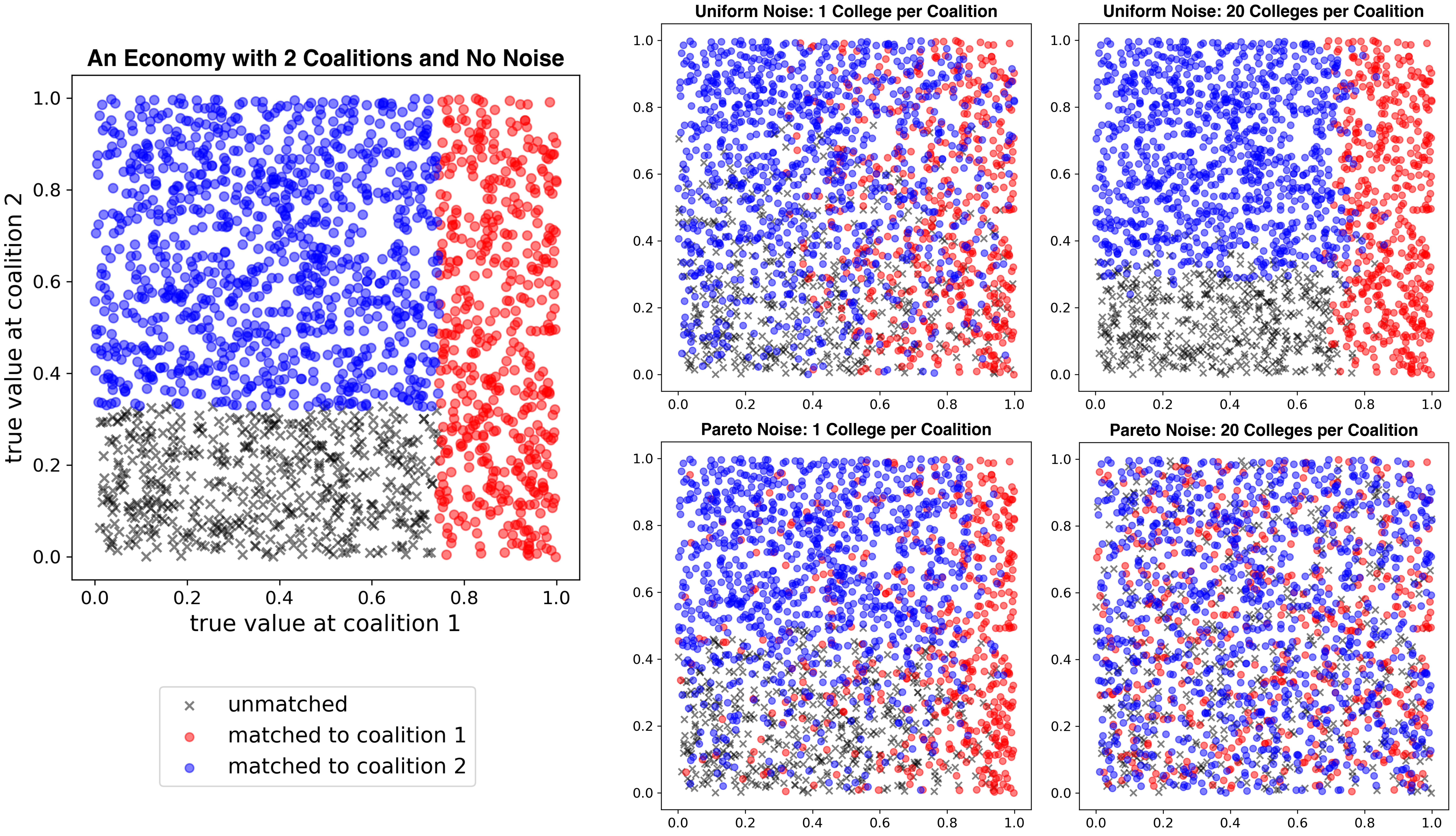
Theorem 2 (Informal)

If noise distribution $\mathscr{D}$ is long-tailed, for all v,

$$\lim_{C \to \infty} \Pr\left[\text{student with true value } v \text{ is matched}\right] = S.$$


Pareto Noise (Long-Tailed)
colleges
$C = 1$
$C = 2$
$C = 5$
$C = 10$
$C = 25$
$C = 100$
probability of matching
student's true value
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

Results roughly extend to *coalitions* of colleges that share true preferences


An Economy with 2 Coalitions and No Noise
true value at coalition 2
true value at coalition 1
0.0
0.2
0.4
0.6
0.8
1.0
unmatched
matched to coalition 1
matched to coalition 2
Uniform Noise: 1 College per Coalition
Uniform Noise: 20 Colleges per Coalition
Pareto Noise: 1 College per Coalition
Pareto Noise: 20 Colleges per Coalition

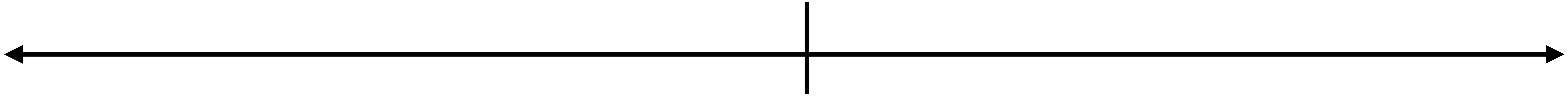
Recap of Theorems

Max-concentrating (light-tailed) noise $\Longrightarrow$ “Wisdom of crowds”

Long-tailed noise $\Longrightarrow$ “Foolishness of crowds”

Intuition

Whether a student matches depends $\approx$ on <u>highest estimated value(s)</u>



$$P_1, P_2, \cdots, P_C = P$$

v matches iff $v + X_c > P_c$ for some $c \in C \iff v + \max_{c \in C} X_c > P$

Proof technique

Given market-clearing cutoffs $P_1, P_2, \cdots, P_C$:

A student with true value v matches iff $v + X_c > P_c$ for some $c \in C$.

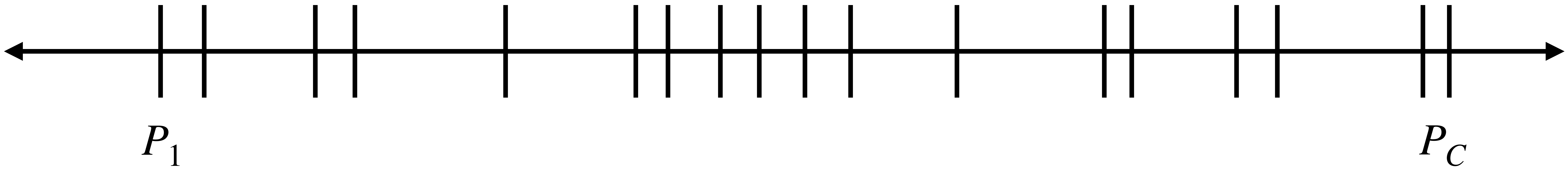
Proof technique

How to determine cutoffs?

Proof technique

Take cutoffs $P_1, P_2, \cdots, P_C$ as given, use market-clearing property as needed

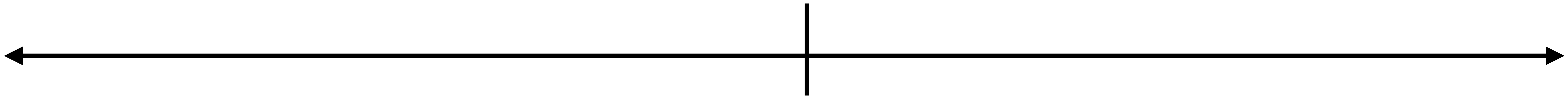
Proof sketch for Theorem 1 (Wisdom of Crowds)



$$P_1 \leq P_2 \leq \cdots \leq P_C$$

“Equal Cutoffs Case”

v matches iff $v + \max_{c \in C} X_c > P$



$P_1, P_2, \cdots, P_C = P$

$v > P - \mathbb{E}[X^{(C)}] + \delta \implies \approx$ always matches

$v < P - \mathbb{E}[X^{(C)}] - \delta \implies \approx$ never has cutoff that exceeds P_{C*}

Lemma (Informal)

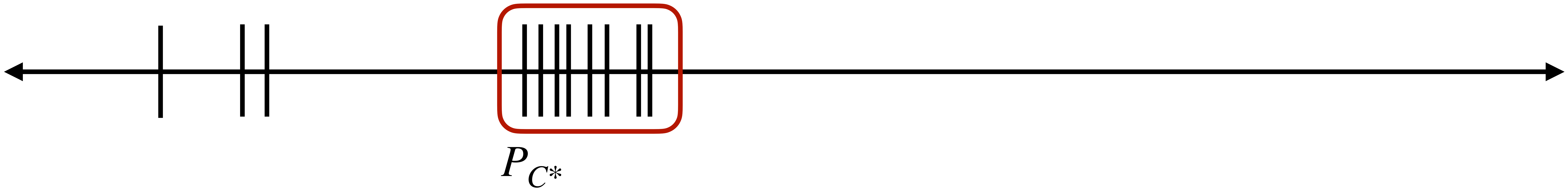
If $C' \subseteq C$ such that $C' \approx C$, then for almost all v,

$$\Pr\left[v \text{ can afford college in } C\right] \approx \Pr\left[v \text{ can afford college in } C'\right].$$

Proof sketch. Suppose otherwise. Then many students can afford a college in $C \backslash C'$ but not C'. Then too many students match to colleges in $C \backslash C'$.

“Early Dense Interval”

Suppose $[P_{C*}, P_{C*} + \delta]$ contains $C/\log C$ cutoffs for C^* “small”

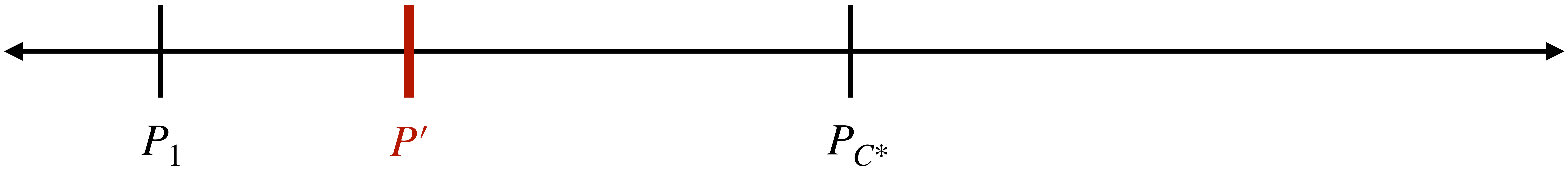


$v > P_{C*} - \mathbb{E}[X^{(C/\log C)}] + 2\delta \implies \approx$ always can afford college in interval

$v < P_{C*} - \mathbb{E}[X^{(C/\log C)}] - \delta \implies \approx$ never has cutoff that exceeds P_{C*}

"Early Big Gap"

Suppose $P_{C*} - P_1$ is "big" for C^* "small"



$v > P' \implies$ almost always can afford P_1

$v < P' \implies$ almost never can afford $P_{C*}, P_{C*+1}, \cdots, P_C$

Pigeonhole principle $\Longrightarrow$ Exists “early dense interval” or “early big gap”

Related Work

- Stable matching with imperfect information
  - Learning-theoretic (Dai and Jordan, 2021; Ionescu et al., 2021; Jagadeesan et al., 2023; Jeloudar et al., 2021; Liu et al., 2020)
  - Game-theoretic (Bikhchandani, 2017; Chakraborty et al., 2010; Chen and Hu, 2020, 2023; Liu, 2020; Liu et al., 2014)
  - Information-theoretic (Ashlagi et al., 2020; Gonczarowski et al., 2019; Immorlica et al., 2020; Kanoria and Saban, 2021; Shi, 2023)
  - Our work: “Statistical”

Related Work

- Random matching markets (Ashlagi et al., 2017; Immorlica and Mahdian, 2003; Pittel, 1989)
- Our work: Consider stable matchings w.r.t. imperfect preferences