Sample Complexity of Auction Design and Optimal Stopping

Zhiyi Huang
The University of Hong Kong

Based on:

Huang, Mansour, Roughgarden (EC 2015)
Huang, Devanur, Psomas (STOC 2016)
Bubeck, Devanur, Huang, Niazadeh (EC 2017)
Guo, Huang, Zhang (STOC 2019)
Guo, Huang, Tang, Zhang (COLT 2021)

Auctions

- 1 item for sale to n bidders
- Each bidder i has a private value v_i **independently** drawn from a prior D_i
- Bidders report values
- Seller picks a winner based on the reported values
- Seller picks a price that the winner pays

Example
Second Price Auction with a Reserve Price

reserve price $1.5

$2

$3

Kevin pays
max($1.5, $2)

$1

Myerson's Solution

Myerson 1981

Optimal auction is characterized by **virtual values**

- For any bidder i, any value v_i, the virtual value is:

$$\phi_i(v_i) = v_i - \frac{1 - F_i(v_i)}{f_i(v_i)}$$

- Bidder with highest non-negative virtual value wins
    - We will omit a caveat called **ironing** in this talk
- Winner pays the **threshold value** above which she wins

Sample Complexity of Optimal Auctions

Cole, Roughgarden 2014

- Assume only sample oracle access to the prior
- Algorithm takes N i.i.d. samples, returns an auction

How many samples are sufficient and necessary to learn an auction that is a $1-\epsilon$ *approximation?*

- Define **approximation ratio** to be

$$\mathbb{E}_{v_1, v_2, \ldots, v_N \sim D}\left[\frac{\text{Rev}(A(v_1, v_2, \ldots, v_N), D)}{\text{Opt}(D)}\right]$$

The learning process is a process of choosing an appropriate function from a given set of functions.

—Vladimir Vapnik

Learning

- **Type space** $\mathscr{T}$
- **Hypothesis space** $\mathscr{H}$; hypothesis $h \in \mathscr{H}$ is a mapping from $\mathscr{T}$ to $[0,1]$
- **Distribution** D over $\mathscr{T}$
- Learn $h \in \mathscr{H}$ via i.i.d. samples from D to minimize or maximize:

$$\mathbb{E}_{t \sim D}\, h(t)$$

Example: Binary Classification

- Type space $\mathscr{T}$: **feature-label pairs**, e.g., $\{(x, y) : x \in \mathbb{R}^n, y = \pm 1\}$
  feature label
- Hypothesis space $\mathscr{H}$: e.g., **linear classifiers**
  - Each $h \in \mathscr{H}$ corresponds to some $a \in \mathbb{R}^n, b \in \mathbb{R}$
  - Let $h(x, y) = 1$ if $\langle a, x \rangle + b$ and y disagree on signs, and 0 otherwise
- Distribution D over $\mathscr{T}$
- Learn $h \in \mathscr{H}$ via i.i.d. samples from D to minimize $\mathbb{E}_{(x,y)\sim D}\, h(x, y)$
  classification error

Sample Complexity

What is the number of samples needed to learn $h \in \mathscr{H}$ up to ϵ-optimal?

Conventional wisdom: decided by **degree of freedom** of $\mathscr{H}$

Classification

If $\mathscr{H}$ has VC-dimension d, $\tilde{\Theta}(d\epsilon^{-2})$ samples are sufficient and necessary.

Vapnik and Chervonenkis 1969, and follow up works

Sample Complexity of Bayesian Optimization Problems

Auctions, Prophet Inequality, and Pandora's Problem

Example: Auctions

Cole and Roughgarden 2014

- Type space $\mathcal{T}$: **value vectors** $v \in [0,1]^n$
- Hypothesis space $\mathcal{H}$: **truthful auctions**
  - Each $h \in \mathcal{H}$ corresponds to a (truthful) auction
  - Let $h(v)$ be the revenue of running the auction on value vector v
- Distribution D over $\mathcal{T}$
- Learn $h \in \mathcal{H}$ via i.i.d. samples from D to maximize $\mathbb{E}_{v \sim D}\, h(v)$

Sample Complexity

- What is the number of samples needed to learn $h \in \mathscr{H}$ up to ϵ-optimal?
- **Conventional wisdom:** decided by **degree of freedom** of $\mathscr{H}$
- Degree of freedom is $\tilde{O}(n\epsilon^{-1})$, unclear if it gives tight sample complexity

Auction: 1 Item, n Buyers

We need at most $\tilde{O}(n\epsilon^{-3})$ samples, and at least $\Omega(\epsilon^{-2})$ samples.

Morgenstern and Roughgarden 2015
Huang, Devanur, and Psomas 2016
Gonczarowski and Nisan 2017
Syrgkanis 2017

Huang, Mansour, and Roughgarden 2015

Example: Prophet Inequality

- Positive rewards $v_i \sim D_i$, $1 \leq i \leq n$, independently
- One reward arrives at a time, must immediate decide to accept it or not
- Aim to maximize reward

Can get at least $0.5\, \mathbb{E} \max_{1 \leq i \leq n} v_i$ in general; for i.i.d. at least $0.745\, \mathbb{E} \max_{1 \leq i \leq n} v_i$.

Krengel, Sucheston, and Garling 1978

Samuel-Cahn 1984

Hill and Kertz 1982

Correa et al. 2017

- Much recent attention, partly due to applications in auction design

Chawla, Hartline, Malec, and Sivan 2010

Example: Prophet Inequality

Correa, Dütting, Fischer, and Schewior 2019

- Type space $\mathcal{T}$: **reward vectors** $r \in [0,1]^n$
- Hypothesis space $\mathcal{H}$: **algorithms for prophet inequality**
  - Each $h \in \mathcal{H}$ corresponds to such an algorithm
  - Let $h(r)$ be the accepted reward by algorithm w.r.t. reward vector r
- Distribution D over $\mathcal{T}$
- Learn $h \in \mathcal{H}$ via i.i.d. samples from D to maximize $\mathbb{E}_{r \sim D}\, h(r)$

Sample Complexity

- What is the number of samples needed to learn $h \in \mathcal{H}$ up to ϵ-optimal?
- **Conventional wisdom:** decided by **degree of freedom** of $\mathcal{H}$
- SOTA uses problem specific arguments

Prophet Inequality

We need at most $\tilde{O}(n^2\epsilon^{-2})$ samples; also at most $\tilde{O}(n\epsilon^{-7})$ samples.

Correa et al. 2019 Rubinstein, Wang, and Weinberg 2020

Example: Pandora's Problem

- n boxes, with independent rewards $r_i \sim D_i$, and cost c_i, $1 \leq i \leq n$
- Algorithm in each step opens another box, or accepts best reward so far
- Aim to maximize reward minus total cost

Optimal strategy has **simple structure**:

1. Set a reserve price for each box
2. Open boxes in descending order of reserve prices
3. Accept first reward exceeding the reserve price

Weitzman 1979 (recently in AGT, see Beyhaghi and Kleinberg 2019; Chawla et al. 2020)

Example: Pandora's Problem

- Type space $\mathscr{T}$: **reward vectors** $r \in [0,1]^n$ (costs are fixed)
- Hypothesis space $\mathscr{H}$: **algorithms for Pandora's problem**
  - Each $h \in \mathscr{H}$ corresponds to such an algorithm
  - Let $h(r)$ be the reward minus cost by algorithm w.r.t. reward vector v
- Distribution D over $\mathscr{T}$
- Learn $h \in \mathscr{H}$ via i.i.d. samples from D to maximize $\mathbb{E}_{r \sim D}\, h(r)$

One theory to rule them all

Even without knowing much about the problems themselves

Theory of Independent Data Dimensions

Part 1: **Bounded**, **finite-support**, product distribution

Part 2: **Bounded** product distribution and **strongly monotone problems**

Part 3: Product distribution and **strongly monotone problems**

Theory of Independent Data Dimensions

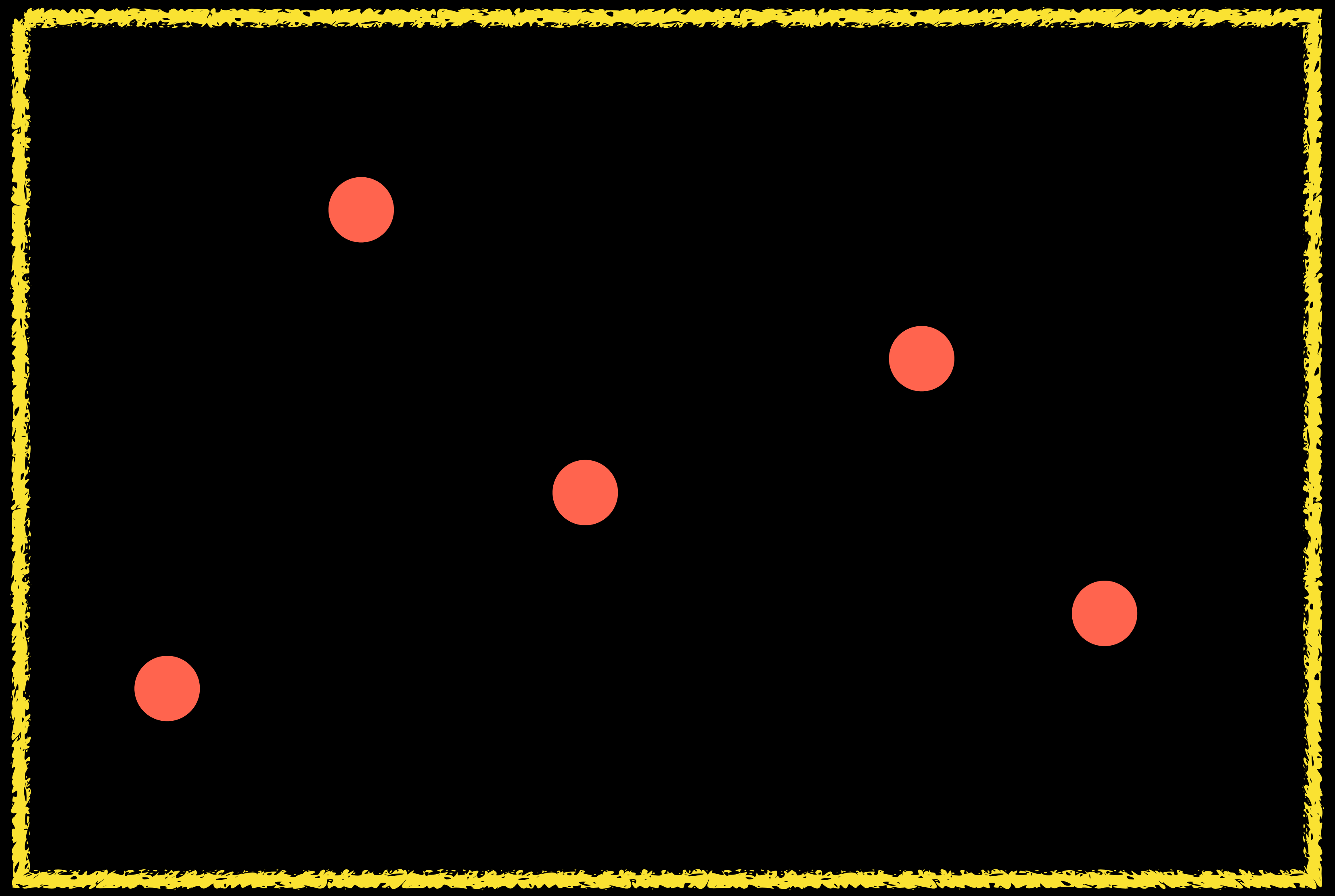
Part 1: If D is a product distribution over n dimensions, each of which has **support size** $\leq k$, then $O(nk\epsilon^{-2}\log\delta^{-1})$ samples can learn an ϵ-optimal $h \in \mathscr{H}$ with probability at least $1-\delta$.

Guo, Huang, Tang, and Zhang 2020
(slightly worse bound implicitly in Gonczarowski and Weinberg 2018)

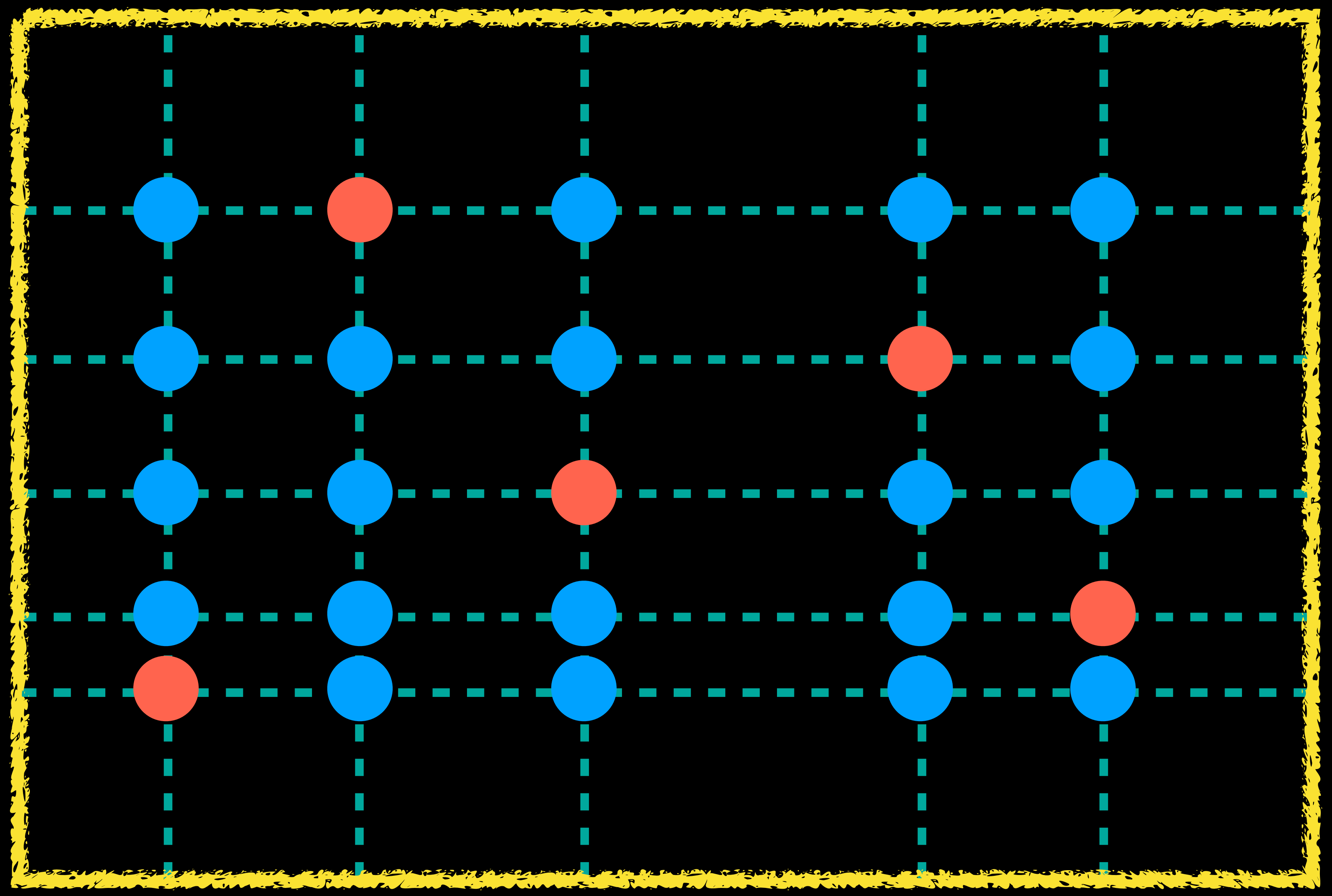
	discretized support size	**sample complexity**	**previous bound**
Auction	ϵ^{-1}	$O(n\epsilon^{-3}\log\delta^{-1})$	$\tilde{O}(n\epsilon^{-3})$
Prophet Inequality	ϵ^{-1}	$O(n\epsilon^{-3}\log\delta^{-1})$	$\tilde{O}(n^2\epsilon^{-2}), \tilde{O}(n\epsilon^{-7})$
Pandora's Problem	ϵ^{-1}	$\tilde{O}(n\epsilon^{-3})$	-

Product Empirical

- $x_1, x_2, \ldots, x_N$ i.i.d. from product distribution $D = \times_{i=1}^{n} D_i$
- Let E_i be uniform over $x_{1i}, x_{2i}, \ldots, x_{Ni}$ from D_i
- Let **product empirical distribution** be $E = \times_{i=1}^{n} E_i$



empirical



product empirical

PERM

Product Empirical Risk Minimizer/Reward Maximizer

- $x_1, x_2, \ldots, x_N$ i.i.d. from product distribution $D = \times_{i=1}^{n} D_i$
- Let E_i be uniform over $x_{1i}, x_{2i}, \ldots, x_{Ni}$ from D_i
- Let **product empirical distribution** be $E = \times_{i=1}^{n} E_i$
- Let **product empirical risk minimizer/reward maximizer** be:

$$\arg\min_{h \in \mathcal{H}} \mathbb{E}_{x \sim E}\, h(x) \quad \text{and} \quad \arg\max_{h \in \mathcal{H}} \mathbb{E}_{x \sim E}\, h(x)$$

Product Distributions are Learnable

If D is a product distribution over n dimensions, each of which has **support size** $\leq k$, the product empirical E from $O(nk\epsilon^{-2}\log\delta^{-1})$ samples satisfies:

$$\text{Hellinger}(D, E) \leq \epsilon$$

Guo, Huang, Tang, and Zhang 2020

$$\text{Hellinger}(D, E)^2 \approx \sum_{i=1}^{n} \sum_{t_i \in \mathcal{T}_i} \left(\sqrt{D_i(t_i)} - \sqrt{E_i(t_i)}\right)^2$$

additivity

Implication: For any $f : \mathcal{T} \mapsto [0,1]$, we have $\mathbb{E}_{t\sim D}\, f(t) \approx_\epsilon \mathbb{E}_{t\sim E}\, f(t)$

Reduction to Vector Concentration

Assuming D is uniform

χ^2-distance

$$\sum_{i,t_i} \left(\sqrt{D_i(t_i)} - \sqrt{E_i(t_i)}\right)^2 = \sum_{i,t_i} \frac{\left(D_i(t_i) - E_i(t_i)\right)^2}{\left(\sqrt{D_i(t_i)} + \sqrt{E_i(t_i)}\right)^2} \le \sum_{i,t_i} \left(\frac{D_i(t_i) - E_i(t_i)}{\sqrt{D_i(t_i)}}\right)^2$$

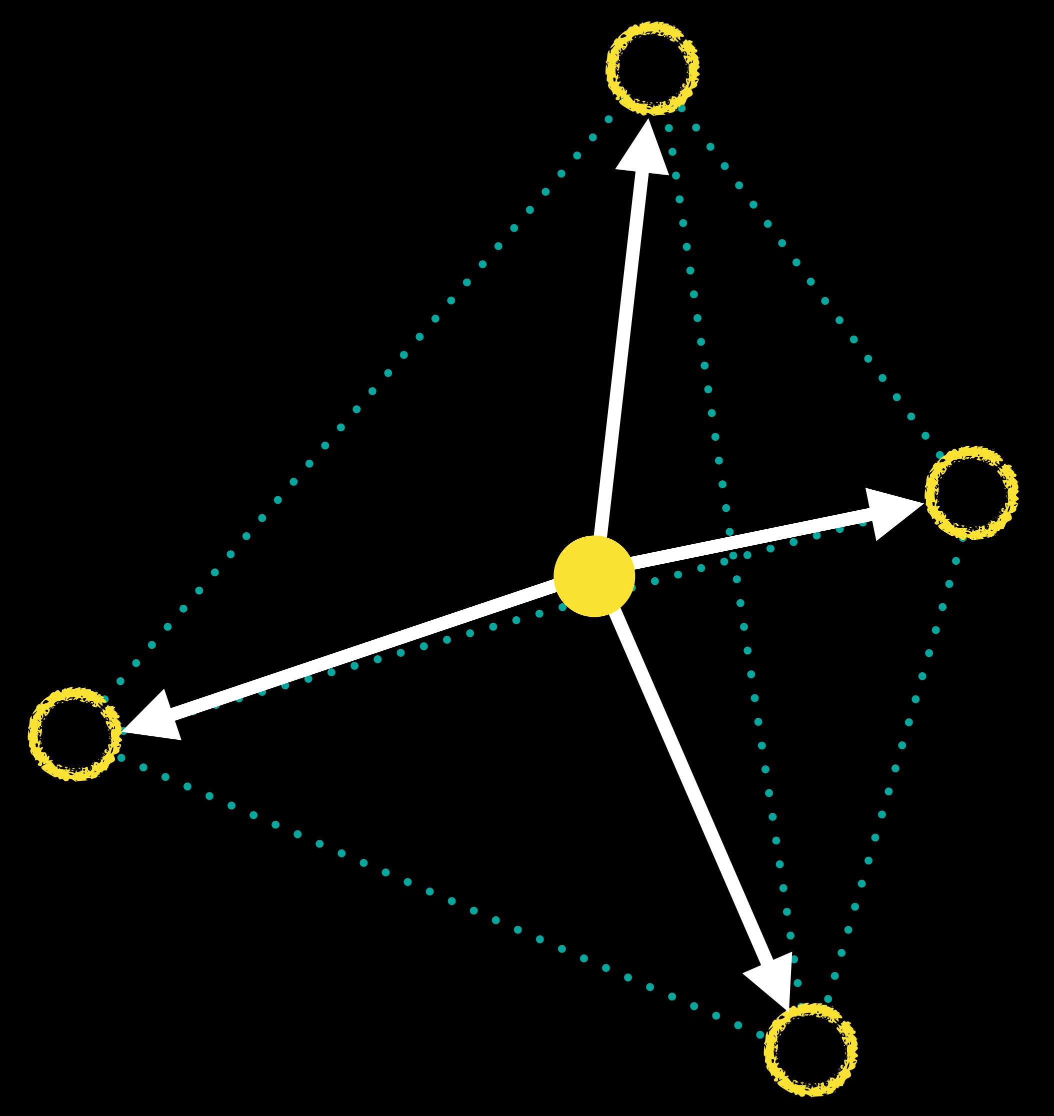
- For each sample x_k, define a random vector v_k with coordinates (i, t_i):

$$v_{k,i,t_i} = \frac{D_i(t_i) - \mathbf{1}(x_{k,i} = t_i)}{\sqrt{D_i(t_i)}}$$

- It remains to show with $N = O(nk\epsilon^{-2}\log\delta^{-1})$ samples:

$$\sum_{i,t_i} \left(\frac{D_i(t_i) - E_i(t_i)}{\sqrt{D_i(t_i)}}\right)^2 = \left\| \frac{1}{N} \sum_{k=1}^{N} v_k \right\|_2^2 \le \epsilon^2$$

average of i.i.d. zero-mean random vectors



$$\left\|\frac{1}{N}\sum_{k=1}^{N} v_k\right\|_2 = \left(\left\|\frac{1}{N}\sum_{k=1}^{N} v_k\right\|_2 - \mathbb{E}\left\|\frac{1}{N}\sum_{k=1}^{N} v_k\right\|_2\right) + \mathbb{E}\left\|\frac{1}{N}\sum_{k=1}^{N} v_k\right\|_2 \leq \epsilon$$

vector concentration inequality

simple exercise
via Cauchy-Schwarz

Chernoff-Hoeffding for Random Vectors

arbitrary norm

For independent v_k s.t. $\mathbb{E}\, v_k = 0$, $\|v_k\| \leq L$, with probability at least $1 - \delta$:

$$\left\|\frac{1}{N}\sum_{k=1}^{N} v_k\right\| - \mathbb{E}\left\|\frac{1}{N}\sum_{k=1}^{N} v_k\right\| \leq L \cdot \sqrt{\frac{\log \delta^{-1}}{N}}$$

zero in original Chernoff-Hoeffding

Ledoux and Talagrand 1991

$$v_{k,i,t_i} = \frac{D_i(t_i) - \mathbf{1}(x_{k,i} = t_i)}{\sqrt{D_i(t_i)}} \Rightarrow L \approx \sqrt{nk}$$

$$N = O(nk\epsilon^{-2} \log \delta^{-1})$$

Theory of Independent Data Dimensions

Part 1: Bounded, finite-support, product distribution

Part 2: **Bounded** product distribution and **strongly monotone problems**

Part 3: Product distribution and **strongly monotone problems**

Theory of Independent Data Dimensions

Part 2: If D is a product distribution over n dimensions, and the class of hypotheses $\mathcal{H}$ is **strongly monotone**, then $\tilde{O}(n\epsilon^{-2})$ samples can learn an ϵ-optimal $h \in \mathcal{H}$ with probability at least $1 - \delta$.

Guo, Huang, and Zhang 2019; Guo, Huang, Tang, and Zhang 2020

stochastic dominance

$\mathcal{H}$ is **strongly monotone** if for any $D' \succeq D$ and the optimal h^* w.r.t. D:

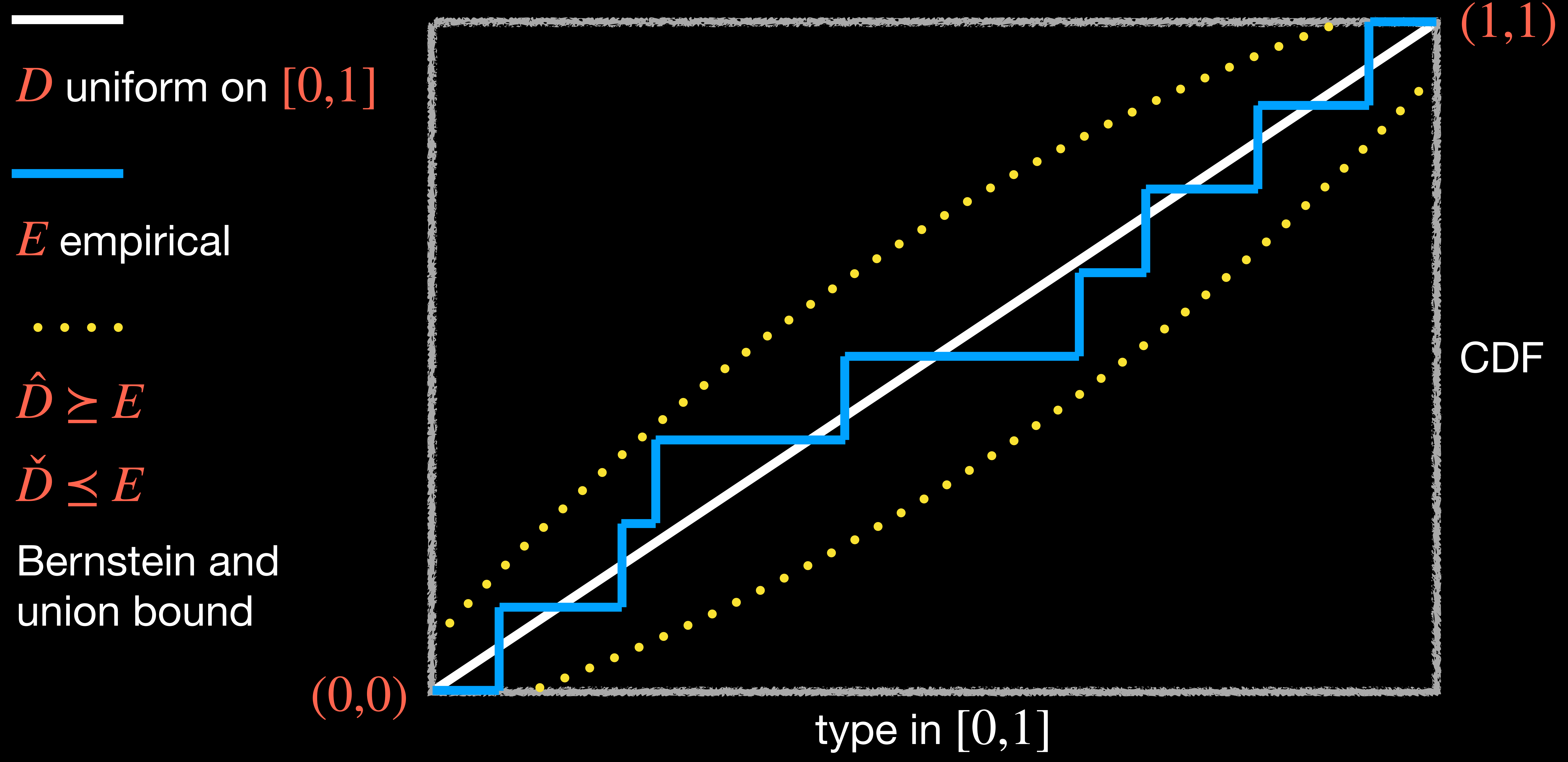
$$\mathbb{E}_{t\sim D'} h^*(t) \geq \mathbb{E}_{t\sim D} h^*(t) = \max_{h\in\mathcal{H}} \mathbb{E}_{t\sim D} h(t)$$

Auctions, Prophet Inequality, and Pandora's Problem are strongly monotone!

matching lower bound

current lower bound $\Omega(n)$

matching lower bound



😔 Hellinger$(D, E) = 1$ for any finite number of samples

😄 Hellinger$(D, \hat{D})$, Hellinger$(D, \check{D}) \leq \epsilon$ with $N = \tilde{O}(n\epsilon^{-2})$ samples

True distribution D

Product empirical E

PERM $h^* = \arg\max_{h \in \mathcal{H}} \mathbb{E}_{t \sim E}\, h(t)$

Auxiliary distributions $\hat{D} \geq E \geq \check{D}$

$$\text{Hellinger}(D, \hat{D}) \leq \epsilon$$

$$\text{Hellinger}(D, \check{D}) \leq \epsilon$$

Goal:

$$\mathbb{E}_{t \sim D}\, h^*(t) \gtrsim_\epsilon \max_{h \in \mathcal{H}} \mathbb{E}_{t \sim D}\, h(t)$$

$$\begin{aligned}
\mathbb{E}_{t \sim D}\, h^*(t) &\overset{\text{Hellinger}(D,\hat{D}) \leq \epsilon}{\approx_\epsilon} \mathbb{E}_{t \sim \hat{D}}\, h^*(t) \\
&\overset{\text{strong monotonicity}}{\geq} \mathbb{E}_{t \sim E}\, h^*(t) \\
&= \max_{h \in \mathcal{H}} \mathbb{E}_{t \sim E}\, h(t) \\
&\overset{\text{strong monotonicity}}{\geq} \max_{h \in \mathcal{H}} \mathbb{E}_{t \sim \check{D}}\, h(t) \\
&\overset{\text{Hellinger}(D,\check{D}) \leq \epsilon}{\approx_\epsilon} \max_{h \in \mathcal{H}} \mathbb{E}_{t \sim D}\, h(t)
\end{aligned}$$

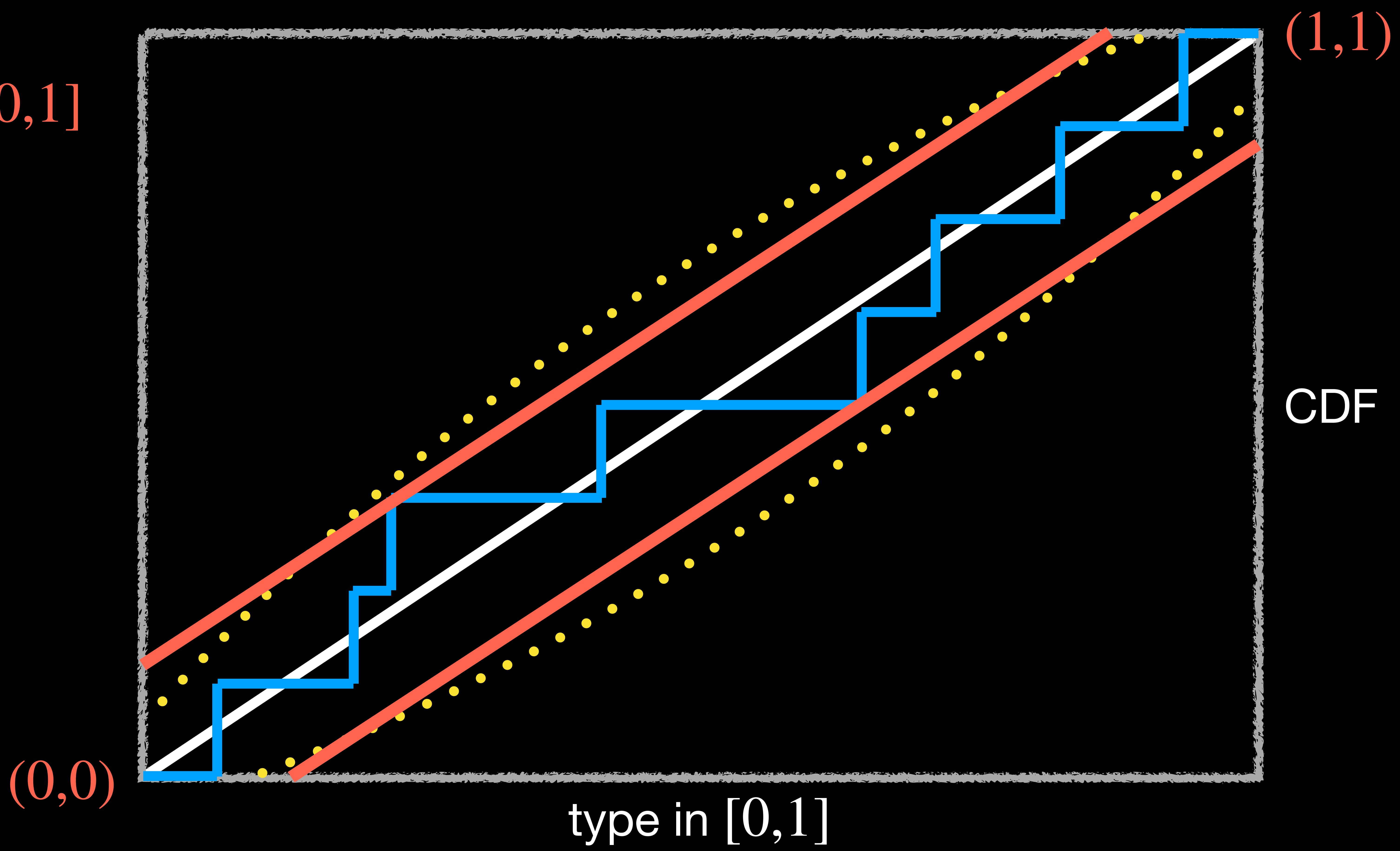
D uniform on $[0,1]$

E empirical

$\hat{D} \succeq E$

$\check{D} \preceq E$

Bernstein and union bound



$\hat{D} \succeq E$

$\check{D} \preceq E$

Dvoretzky-Kiefer-Wolfowitz

DKW needs $N = O(n^2\epsilon^{-2}\log\frac{1}{\delta})$ for Hellinger$(D, \hat{D})$, Hellinger$(D, \check{D}) \leq \epsilon$

quadratic in n better log factor

Open Question: Is there a Bernstein-style DKW inequality?

$$\forall x \in [0,1] : \quad |F_D(x) - F_E(x)| \leq$$

$$\sqrt{\frac{F_D(x)(1 - F_D(x)) \log N/\delta}{N}}$$

Bernstein and Union Bound

$$\sqrt{\frac{\log 1/\delta}{N}}$$

Dvoretzky-Kiefer-Wolfowitz

$$\sqrt{\frac{F_D(x)(1 - F_D(x)) \log 1/\delta}{N}}$$

Bernstein-style Dvoretzky-Kiefer-Wolfowitz (?)

Theory of Independent Data Dimensions

Part 1: Bounded, finite-support, product distribution

Part 2: Bounded product distribution and strongly monotone problems

Part 3: Product distribution and **strongly monotone problems**

Theory of Independent Data Dimensions

Part 3: We can learn a **dominated product empirical** $\tilde{E} \preceq D$ using $N = \tilde{O}(n\epsilon^2)$ samples such that there is $\tilde{D} \preceq \tilde{E}$ satisfying:

$$\text{Hellinger}(D, \tilde{D}) \leq \epsilon$$

Guo, Huang, and Zhang 2019; Guo, Huang, Tang, and Zhang 2020

Handle **unbounded distributions** in Auctions, Prophet Inequality

One-sided error, using the same number of samples as product empirical

You get what you expect: If $\tilde{h} = \arg\max_{h \in \mathcal{H}} \mathbb{E}_{t \sim \tilde{E}}\, h(t)$, then $\mathbb{E}_{t \sim D}\, \tilde{h}(t) \geq \mathbb{E}_{t \sim \tilde{E}}\, \tilde{h}(t)$

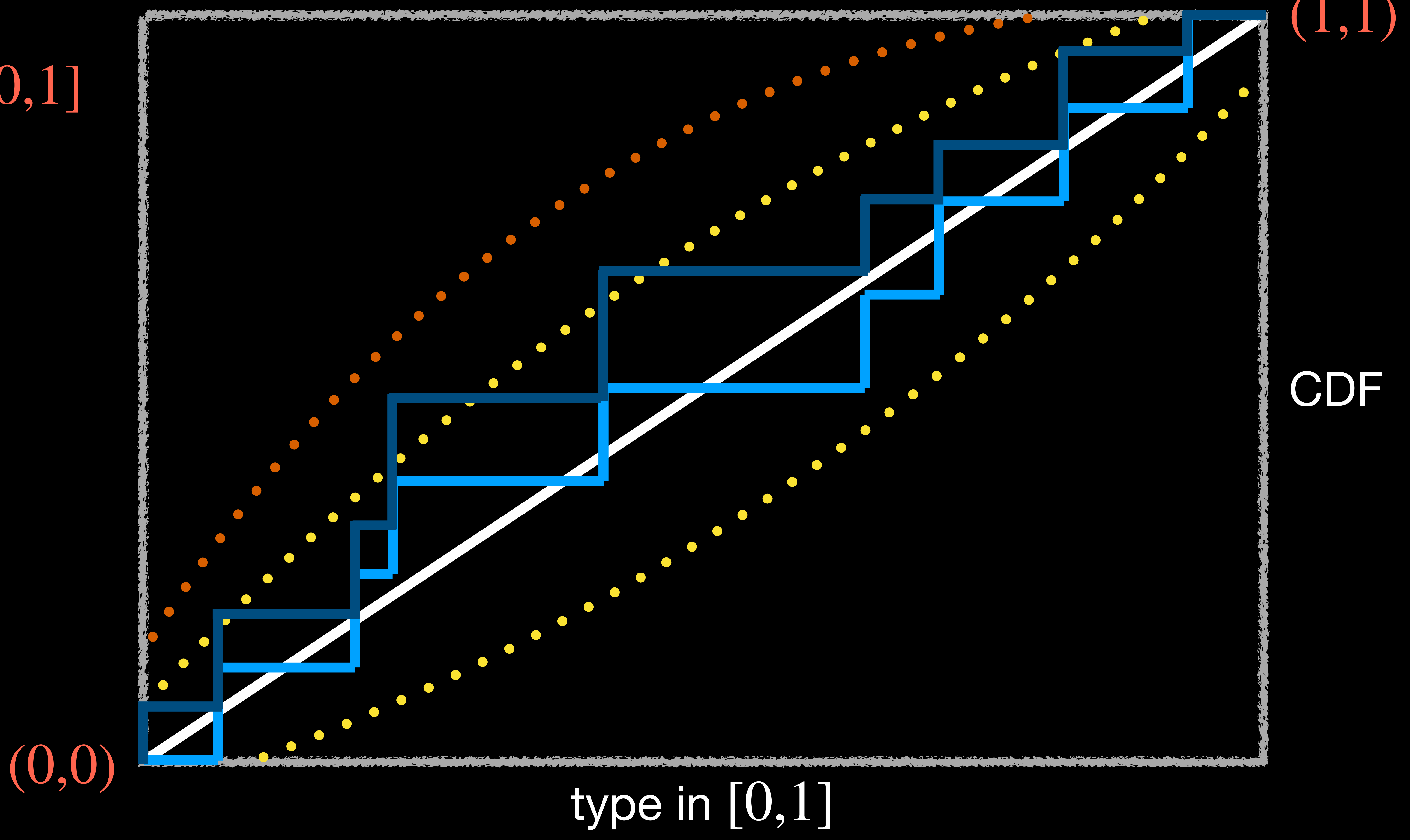
D uniform on $[0,1]$

E empirical

$\hat{D} \succeq E$

$\check{D} \preceq E$

Bernstein and union bound



$\tilde{E}$ dominated empirical

i.e., empirical minus Bernstein error term

$\tilde{D} \preceq E$

i.e., doubled Bernstein error term

Summary

Theory of Independent Data Dimensions

Part 1: Bounded, finite-support, product distribution

- Most general, give best-to-date sample complexity for multi-item auctions

Part 2: Bounded product distribution and **strongly monotone problems**

- Nearly optimal sample complexity for single-item auctions, prophet inequality, Pandora's problem

Part 3: Product distribution and **strongly monotone problems**

- Nearly optimal sample complexity for single-item auctions, prophet inequality with unbounded distributions

Next Step

Structured Correlated Distributions

- Learning **high dimensional** **arbitrarily correlated** distributions is hard
- How about correlated distributions with structures? Recent works in:
    - Auctions
      Brustle, Cai, and Daskalakis 2020
    - Pandora's problem
      Chawla, Gergatsouli, Teng, Tzamos, and Zhang 2020
- **Can we learn a data representation that is independently distributed?**

Next Step

Problem Structure Other than Monotonicity

- Multi-item auction does not even satisfy weak monotonicity
  i.e., optimal revenue could decrease as value distributions get "bigger"
- We believe the current sample complexity is suboptimal
- **Is there another structural property that could help?**

Next Step

Computational Complexity

- Bayesian optimization problems are mostly studied assuming independence
  - Single-item auction
    Myerson 1981
  - Prophet inequality
    Krengel, Sucheston, and Garling 1978
  - Pandora's problem
    Weitzman 1979
- Some mathematical-program-based algorithms require small-support
  - Multi-item auctions
    Cai, and Daskalakis, and Weinberg 2012, 2013
- **Compuational complexity of ERM vs. PERM?**

Thank you!

Questions?