Robust (MD + ML) = *Learned Mechanisms*

Johannes Brustle
LSE

Yang Cai
Yale

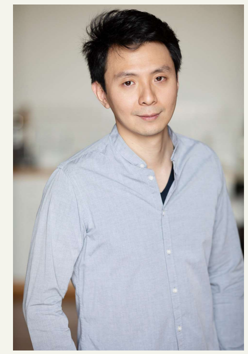
Costis Daskalakis
MIT

1. [BCD] Multi-Item Mechanisms without Item-Independence: Learnability via Robustness. (EC ’20)
2. [CD] Recommender Systems meet Mechanism Design. (EC ’22)

Motivation

what is this
rock and how
should I sell it

private
value v



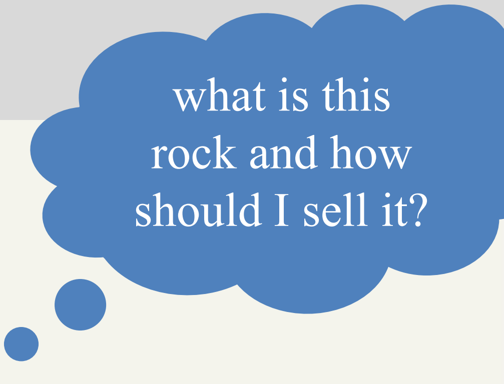
- How to sell an item to optimize revenue?
  - without information about the buyer's value, no meaningful optimization of revenue can be attained.
- ***Bayesian assumption:*** the seller knows a distribution F s.t. $v \sim F$.
  - Private value: We know F, but not the the sampled value v.
  - Quasi-linear Utility: $v \cdot x - p$ if wins the item with prob. x and pays price p.
- [Riley-Zeckhauser'81, Myerson'81]: The optimal mechanism is a take-it-or-leave-it offer of the item at price: $p^* \in \arg\max\{z \cdot (1 - F(z))\}$.

Motivation (cont’d)

v_1

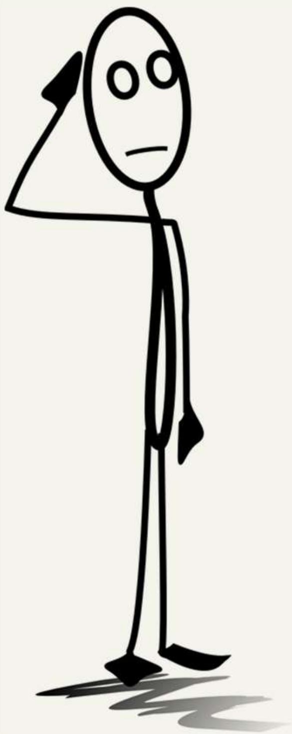
$\vdots$

v_n



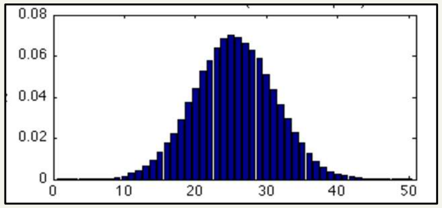
- How to sell an item to optimize revenue?
- [Myerson’81]: When $v_1, \ldots, v_n$ are i.i.d. $\sim F$ optimal auction is second price auction with reserve price $p^* \in \arg\max\{z \cdot (1 - F(z))\}$.
  - similar characterizations of optimal auction if the v_i’s are just independent
- Workhorse in theory and practice of auctions.

Motivation (cont'd)



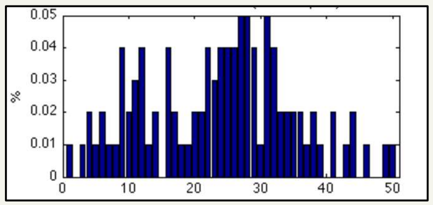
- Where exactly does the prior F come from?
  - **A:** from market research or observation of bidder behavior in se
    the same kind of items in some prior auction + econometric ana
- Hmmm, so our best bet is that we know $\hat{F} \approx F$
- Using $\hat{F}$ instead of F is usually a bad idea
  - overfitting to details of $\hat{F}$
- **What to do?**

Motivation (cont'd)



true F

observations (somewhere)



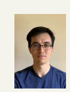
empirical dist'n

econometrics

approximate $\hat{F}$

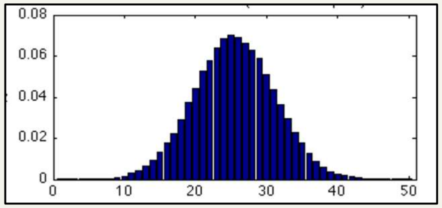
MD

overfitting



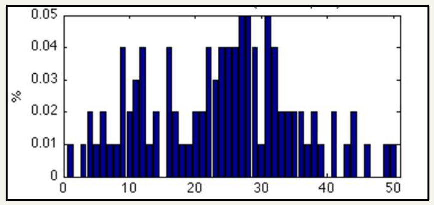
auction

Motivation (cont'd)



true F

observations (somewhere)



empirical dist'n

econometrics

approximate $\hat{F}$

Robust MD

Sample-based MD

-Roughgarden'14, Mohri-Medina'14, Huang et al'14, Morgenstern-
5, Devanur et al'16, Roughgarden-Schrijvers'16, Gonczarowski-Nisan'16,
16, Syrgkanis'17, Cai-Daskalakis'17, Gonczarowski-Weinberg'18, Balcan
t al. '19…]

Robust Mechanism Design

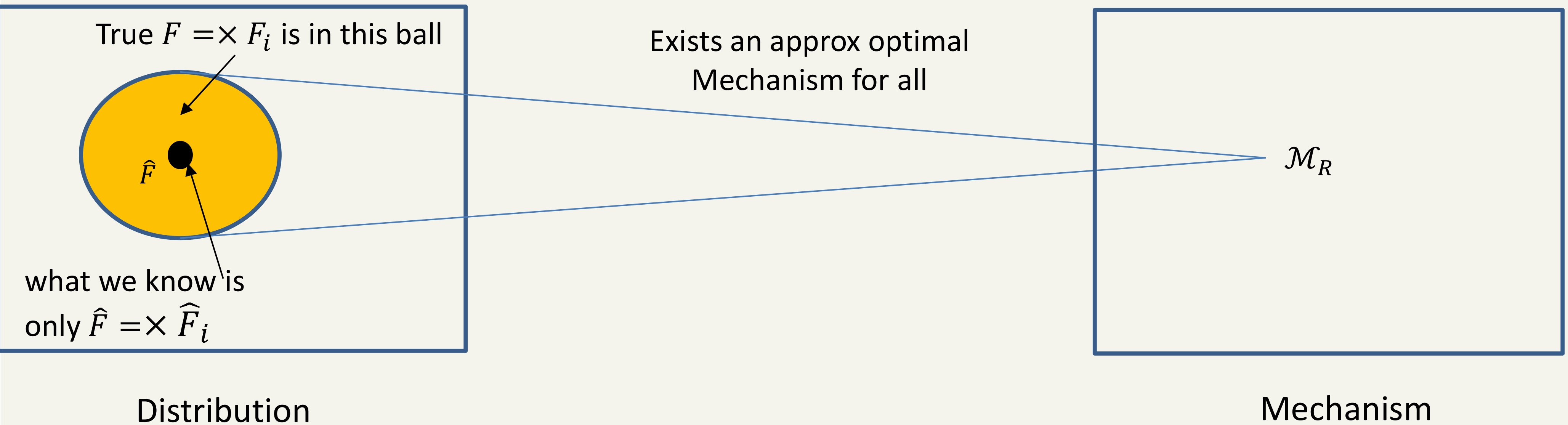
Robust Mechanism Design – One-Item Many-Bidd

ng: n bidders, 1 item, independent values drawn from $F = F_1 \times \cdots \times F_n$, $\forall i$ know $\hat{F}_i$ such that $\mathrm{d}(\hat{F}$

Goal: given $\hat{F}_1, \ldots, \hat{F}_n$ and with no knowledge of $F_1, \ldots, F_n$, find mechanism $\mathcal{M}_R$ such that:

$$\mathrm{Rev}_{\mathcal{M}_\mathrm{R}}(\times_i F_i) \geq \mathrm{OPT}(\times_i F_i) - \mathrm{err}(\epsilon, n)$$

where $\mathrm{err}(\epsilon, n) \longrightarrow 0$ as $\epsilon \longrightarrow 0$.



A priori unclear if such $\mathcal{M}_R$ exists.

Robust Mechanism Design – One-Item Many-Bidd

ng: n bidders 1 item, values drawn from $F = F_1 \times \cdots \times F_n$ for all i, know $\hat{F}_i$ such that $\mathrm{d}(\hat{F}_i, F_i) \le \epsilon$

given $\hat{F}_1, \dots, \hat{F}_n$ and with no knowledge of $F_1, \dots, F_n$, find mechanism $\mathcal{M}_R$ such that:

$$\mathrm{Rev}_{\mathcal{M}_\mathrm{R}}(\times_i F_i) \ge \mathrm{OPT}(\times_i F_i) - \mathrm{err}(\epsilon, n)$$

where $\mathrm{err}(\epsilon, n) \longrightarrow 0$ as $\epsilon \longrightarrow 0$.

Individual Rationality (IR): the buyer has non-negative utility if report truthfully
Dominant Strategy Incentive Compatible (DSIC): reporting truthfully is a dominant strategy

tle-Cai-Daskalakis EC'20]:

Distance d		**ty**
Kolmogorov	$\mathrm{Rev}(\mathcal{M}_R, F) \ge \mathrm{OPT}(\ldots) \ldots (n\epsilon)$ $\mathcal{M}_R$ is IR and DSIC	$\lvert OPT(\hat{F}) - OPT(F)\rvert \le O(n\epsilon)$
Lévy	same	same

$(F, G) = \sup_x |F(x) - G(x)|$

$d_L(F, G) = \inf\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \le G(x) \le F(x + \varepsilon) + \varepsilon, \forall x$

retzky-Kiefer-Wolfowitz Inequality]: With prob. $1 - \delta$, the empirical distribution $O(\frac{\log\ \ /\delta}{\epsilon^2})$ samples is within ϵ in Kolmogorov distance to the original one.

minance.

Multi-Dim Revenue Maximization

⋮

how should I auction seats in all Hong Kong restaurants?

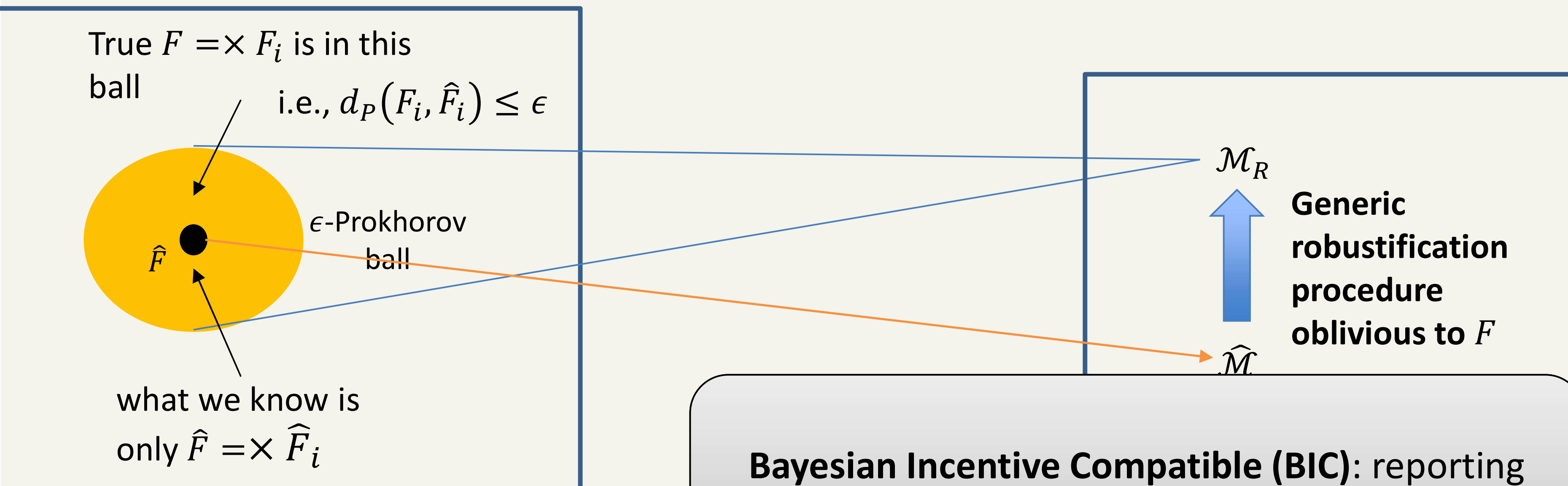
- characterization of revenue optimal mechanism much more challenging and no general characterization is known.[Rochet'85], [Laffont-Maskin-Rochet'87], [McAfee-McMillan'88], [Wilson'93], [Armstrong'96], [Rochet-Chone'98], [Armstrong'99], [Zheng'00], [Basov'01], [Kazumori'01], [Thanassoulis'04], [Vincent-Manelli '06,'07], [Figalli-Kim-McCann'10], [Pavlov'11], [Hart-Nisan'14], [Hart-Reny'15], [Daskalakis-Deckelbaum-Tzamos '17], [Frongillo-Kash '16] …
- Lots of recent progress on various fronts (characterizations, simple-vs-optimal results,…)

Example: additive valuation, $t = (t_1, \dots, t_m)$, $x = (x_1, \dots, x_m)$ and $v_i(t; x) = \sum_{i \in [m]} t_i x_i$

neral Setting: n bidders, multi-dim typ

$$u_i(t; x, p) = v_i(t; x) - p$$

$v_i(t; x) \in [0,1]$, 1-Lipschitz w.r.t. t in ℓ_1 for every allocation x.



Distributions

Bayesian Incentive Compatible (BIC): reporting truthfully is a Bayes-Nash equilibrium

ustness holds for **arbitrary mechanism**

stle-Cai-Daskalakis EC'20]: given a mechanism $\widehat{\mathcal{M}}$ BIC and IR w.r.t. $\hat{F}$, can turn it into robust $\mathcal{M}_R$ su

or all F in the ϵ-ball: $\mathcal{M}_R$ is $\mathrm{appx}(\epsilon, n, m)$- BIC, exactly-IR, and $\mathrm{Rev}_{\mathcal{M}_R}(F) \ge \mathrm{Rev}_{\widehat{\mathcal{M}}}(\hat{F}) - \mathrm{err}(\epsilon, n,$

mplies that $\mathrm{OPT}(F) \approx \mathrm{OPT}(\hat{F})$, so if $\widehat{\mathcal{M}}$ is approx. optimal for $\hat{F}$, $\mathcal{M}_R$ is approx. optimal for F.

Prokohorov Distance

- Prokhorov Distance:
  - Widely used in robust statistical decision theory **[Huber '81, Hampel et al. '86]**.
  - Strassen's Characterization of the Prokhorov distance:

$$d_P(F, \hat{F}) \leq \varepsilon \Leftrightarrow \exists \text{ coupling } \gamma(x, y) \text{ of } F, \hat{F} \text{ s.t. } \Pr_{\gamma}[\|x - y\| > \varepsilon] \leq \varepsilon$$

  - i.e. can couple $F, \hat{F}$ so that, w/ probability $\geq 1 - \varepsilon$, samples are within ε

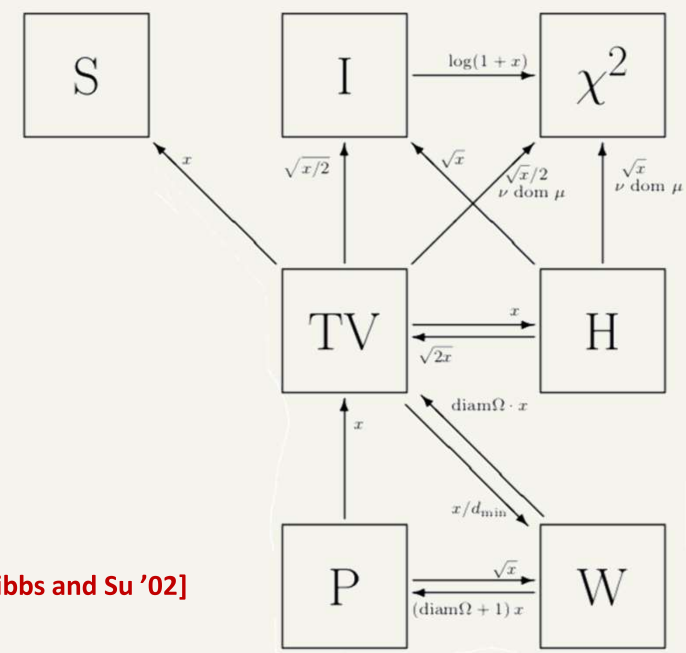


Figure credit: **[Gibbs and Su '02]**

Mechanism Robustness
& Optimal Revenue Continuity

etting: n bidders, quasi-linear utilities, independent multi-dim types in $\mathbb{R}^m$ drawn from $= F_1 \times \cdots \times F_n$, for all i, know $\hat{F}_i$ such that $\mathrm{d_P}(\hat{F}_i, F_i) \leq \epsilon$

Brustle-Cai-Daskalakis EC'20]: Given $\widehat{\mathcal{M}}$ that is BIC and IR w.r.t. $\hat{F}$ construct robust $\mathcal{M}_R$ s.t.

Robustness	Continuity
$\mathrm{Rev}(\mathcal{M}_R, F) \geq \mathrm{Rev}\left(\widehat{\mathcal{M}}, \hat{F}\right) - O(n\eta + nm\sqrt{\eta})$ $\mathcal{M}_R$ is IR and η-BIC ($\eta = nm\epsilon + m\sqrt{n\epsilon}$)	$\left\|OPT\left(\hat{F}\right) - OPT(F)\right\| \leq O(n\eta + nm\sqrt{\eta})$

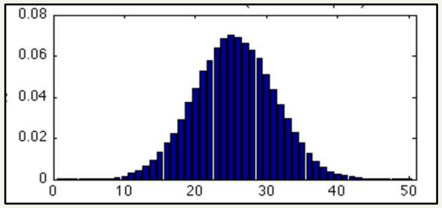
orollary: Given $\widehat{\mathcal{M}}$ that is the revenue-optimal BIC and IR w.r.t. $\hat{F}$, can construct $\mathcal{M}_R$ s.t.

$$\mathrm{Rev}(\mathcal{M}_R, F) \geq \mathrm{OPT}(F) - O(n\eta + nm\sqrt{\eta})$$

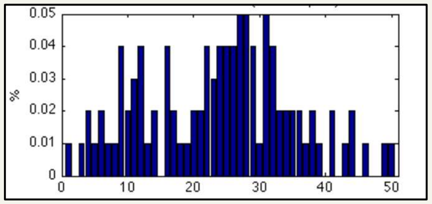
uestion: Can we make $\mathcal{M}_R$ be **exactly BIC** and appx-optimal for all dist'ns in the ball?

- **Yes** if m or $n = 1$; **No** for multiple items multiple bidders (follows from **[Lopomo-Rigotti-Shanno** **+[Tang-Wang'16]**).

Corollary: Modularity (1+1=3)



true F



observations (somewhere)

1

econometrics

approximate $\hat{F}$

Robust MD

2

MD

Econometrics, Stat, ML: solve 1, Prokhorov guarantees

+

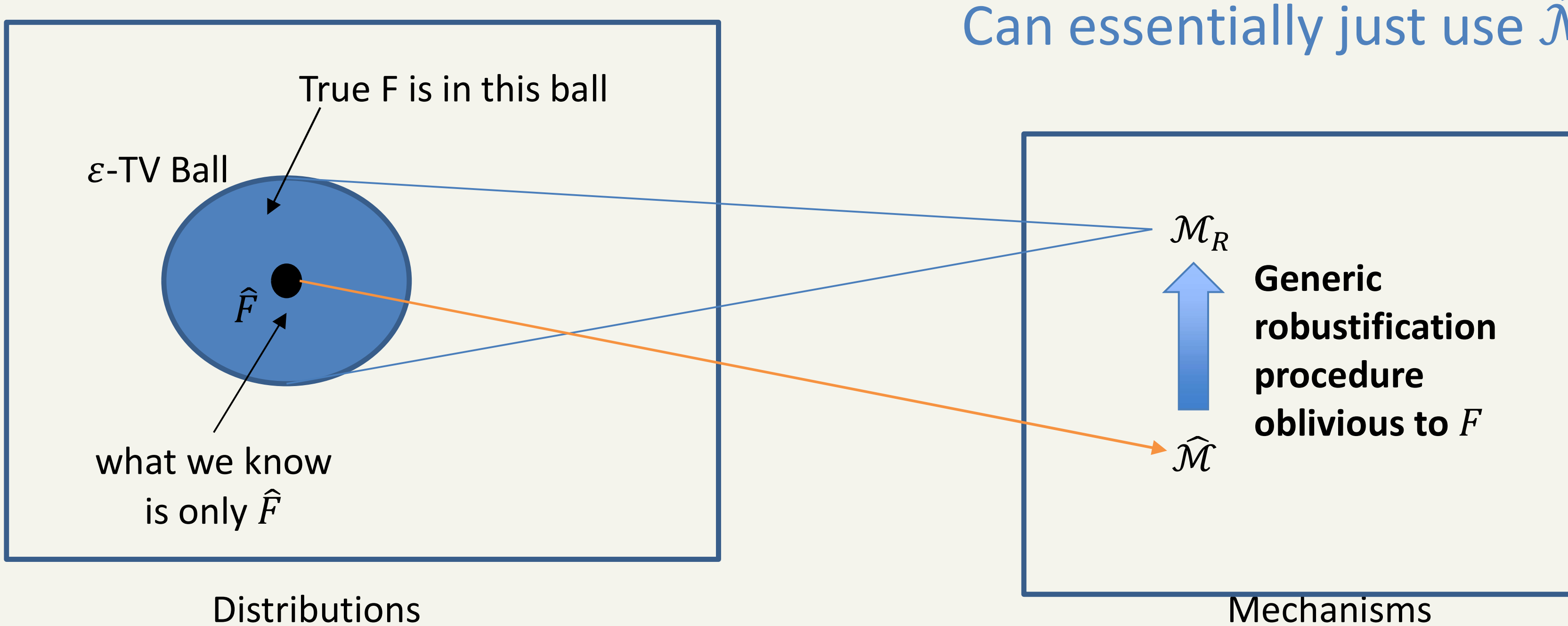
MD: solve 2 for $\hat{F}$

Approximately Optimal Mechanism for unknown F

Proof Vignettes of Robustness

TV-Robustness

Can essentially just use $\widehat{\mathcal{M}}$!

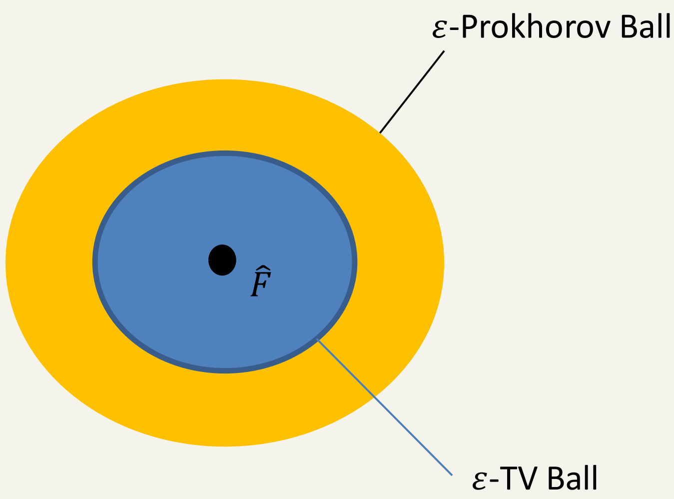


Total variation distance: $d_{TV}(F, \hat{F}) = \sup_{event\ \mathcal{E}} |F(\mathcal{E}) - \hat{F}(\mathcal{E})|$,

$d_{TV}(F, \hat{F}) \leq \varepsilon \Leftrightarrow \exists$ coupling $\gamma(x, y)$ of $F, \hat{F}$ s.t. $\Pr_{\gamma}[x \neq y] \leq \varepsilon$

Prokhorov Robustness <=> TV Robustness

- Prokhorov Robustness => TV Robustness

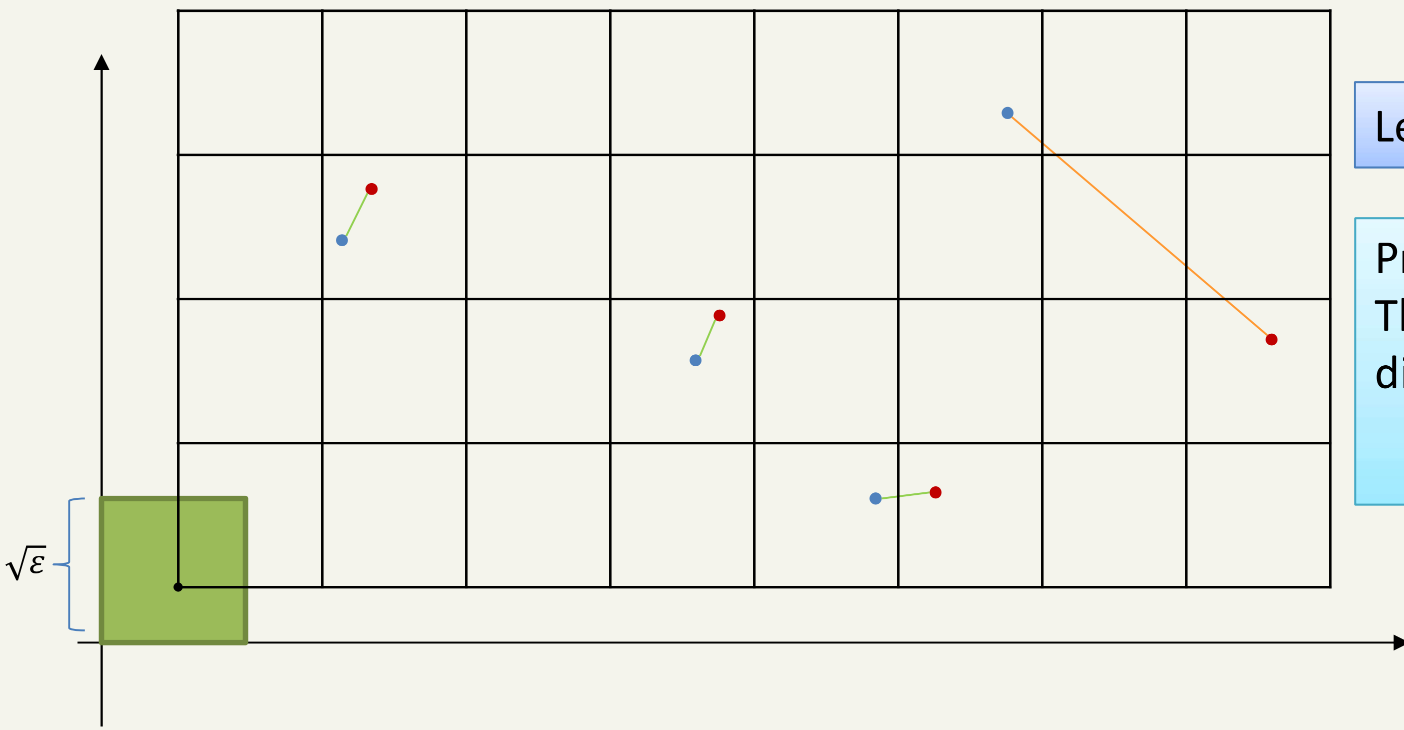


- TV Robustness => Prokhorov Robustness
  - hope a $\text{poly}(n, m, \varepsilon)$-TV Ball contains the ε-Prokhorov Ball
  - but even the $(1-\delta)$-TV Ball does not contain the ε-Prokhorov Ball for arbitrarily small δ.

TV Robustness => Prokhorov Robustness

- Key Idea: round the distributions down to a **random grid**.

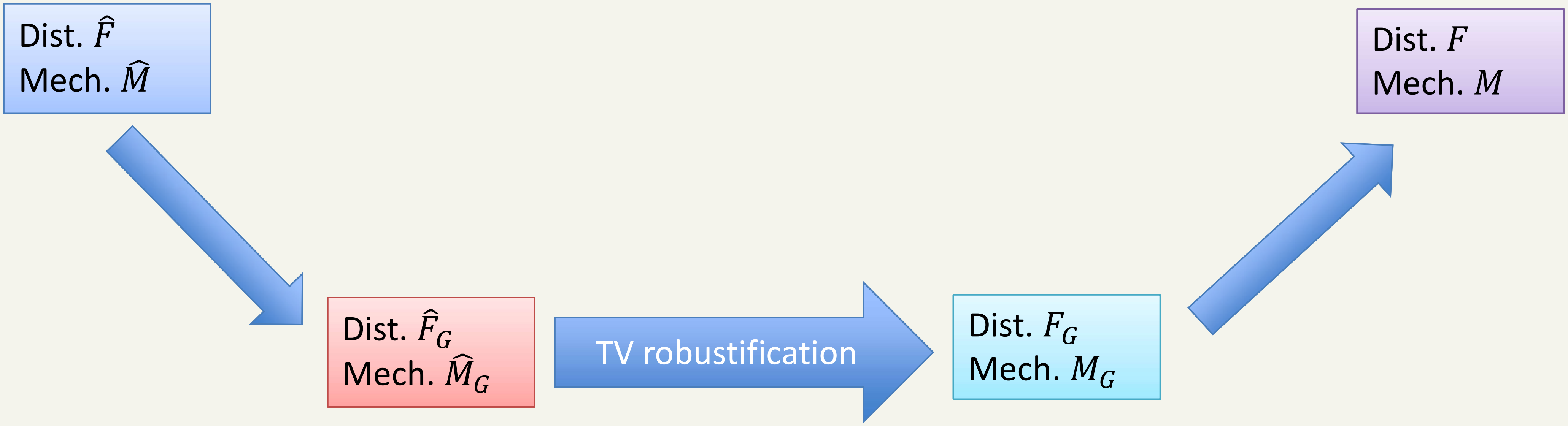
> Definition of the Prokhorov distance: $d_P(F, \hat{F}) \le \varepsilon \Leftrightarrow \exists$ coupling $\gamma(x, y)$,
> s.t. $\Pr_{\gamma}[\|x - y\|_1 > \varepsilon] \le \varepsilon$



> Lemma: $\mathbb{E}[d_{TV}(F_G, \hat{F}_G)] = O($

> Proof: Suppose $\|x - y\|_1 \le \varepsilon$
> The prob. that x and y fall int
> different cubes is exactly:
> $$\sum_{i=1}^{m} \frac{|x_i - y_i|}{\sqrt{\varepsilon}} \le \sqrt{\varepsilon}.$$

Prokhorov Robustification



Sample a random grid G.

Create $\hat{F}_G$ and $\widehat{M}_G$ that is appx-BIC and IR wrt $\hat{F}_G$.

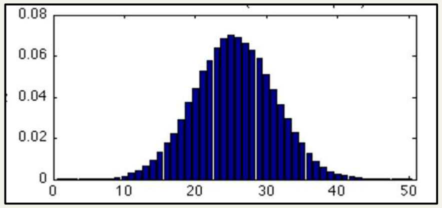
- For any t_i, sample b_i from $\hat{F}$ conditioned on being in the right box, i.e., b_i will be rounded to be in grid G.
- Feed $(b_1, \ldots, b_n)$ to M.

Use TV robustification to obtain M_G that is appx-BIC and IR wrt F_G.

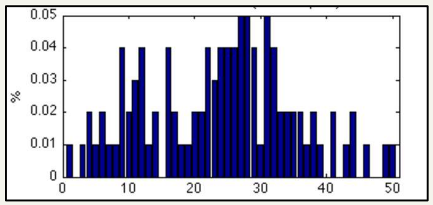
Create M that is appx-BIC and IR wrt F.

- For any t_i, rounded to grid G, and report the rounded type to M_G .

Corollary: Modularity (1+1=3)



true F



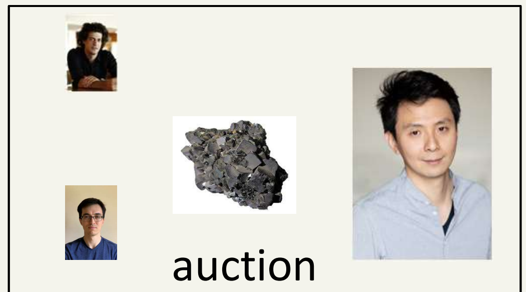
observations (somewhere)

1

econometrics

approximate $\hat{F}$

Robust MD



2

MD

Econometrics, Stat, ML: solve 1, Prokhorov guarantees

+

MD: solve 2 for $\hat{F}$

Approximately Optimal Mechanism for unknown F

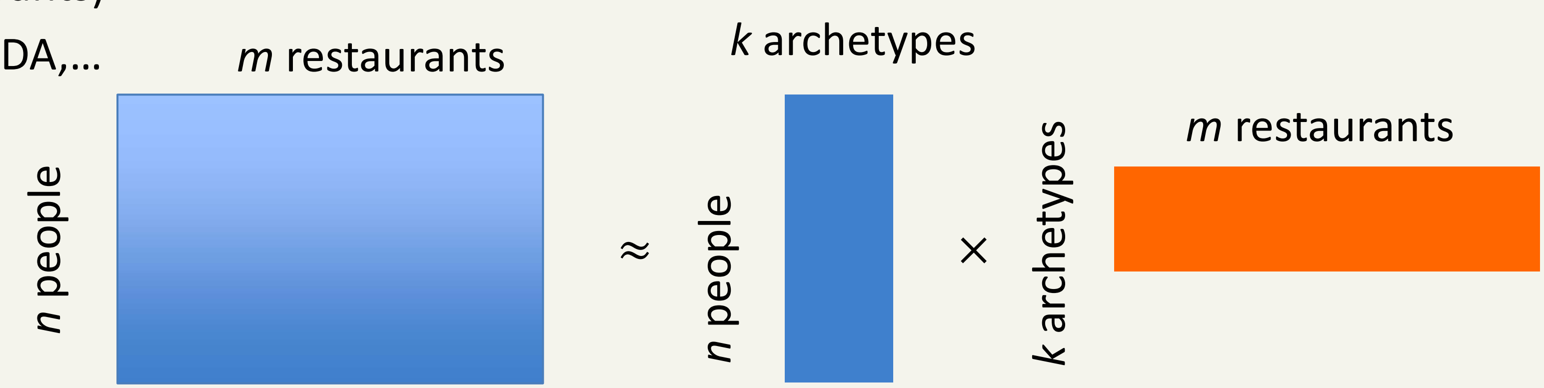
Robust MD meets ML

Example 1: Mechanism Design meets Topic Models

⋮

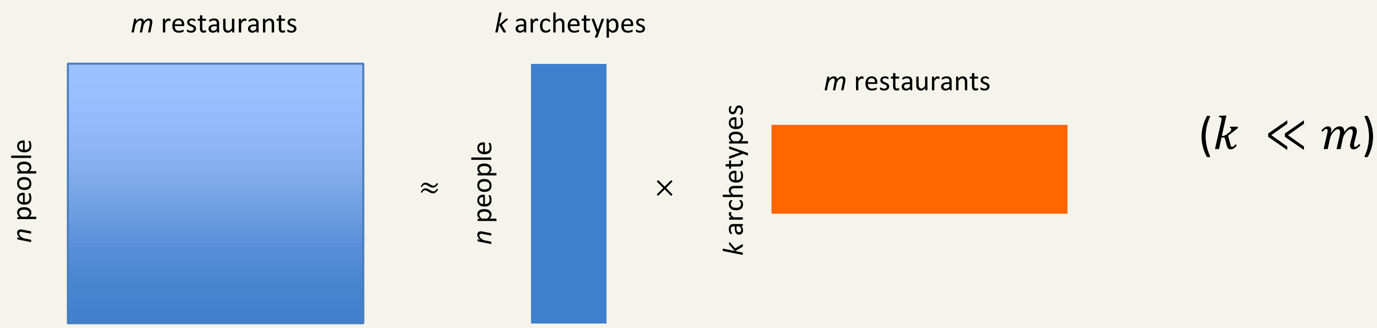
how should I auction seats in all Athens restaurants?

- Topic Models: practically useful (family of) statistical models for high-dimensional data with structure.
- Basic premise: high-dimensional vector $t \in \mathbb{R}^m$ (e.g. m=#restaurants) generated by
  - first sampling a mixture over k archetypes (e.g. food connoisseurs)
  - then outputting m-dimensional vector by combining -- in some way dependent on sampled mixture dimensional samples corresponding to each archetype (e.g. preferences of food connoisseurs for restaurants)
- E.g. NMF, LDA,…



Example 1: Mechanism Design meets Topic Models (cont'd)

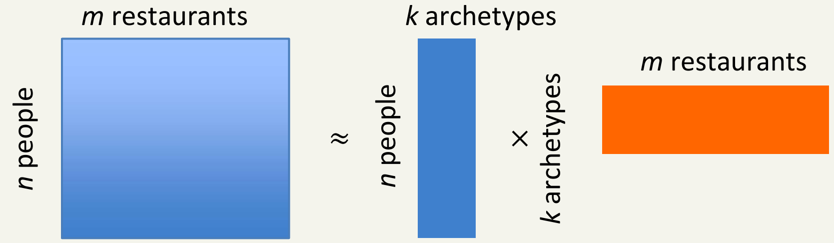
- ***Challenge***: Suppose topic model is good approximation of high-dimensional type distribution F; design good mechanism for F.



- ***Issue***: topic model is only an approximation of reality (true types are close to manifold spanned by topic model samples)
- ***Extra challenge***: impractical to ask bidders to communicate their m-dimensional type
  - how about asking them about their mixture over archetypes?
  - ***Issue 2:*** bidders don't know anything about archetypes!

Example 1: Mechanism Design meets Topic Models (cont'd)

- *Challenge 1*: Suppose topic model is good approximation of high-dimensional type distribution F; design good mechanism for F.



$(k \ll m)$

- *Challenge 2*: impractical to ask bidders to communicate their m-dimensional type, but bidders don't know anything about archetypes!
- 1+1=3 approach:
  - step 1 (ML): ask ML team to learn topic model $\hat{F}$ approximating true F in Prokhorov
  - step 2 (MD): ask MD team to design mechanism $\widehat{\mathcal{M}}$ for topic model $\hat{F}$
    - done right, effective dimensionality is k=#archetypes (rather than m=#restaurants)
    - e.g. (fake) $\widehat{\mathcal{M}}$ can ask bidders for their mixture over archetypes rather than their m-dimensional types
  - step 3 (Robust MD): massage $\widehat{\mathcal{M}}$ into $\mathcal{M}_R$ attaining approximately same revenue on F as $\widehat{\mathcal{M}}$ on $\hat{F}$
    - if $\widehat{\mathcal{M}}$ is α-optimal for $\hat{F}$, then $\mathcal{M}_R$ is $\alpha(ish)$-optimal for F
    - $\mathcal{M}_R$ can be made to ask sparse queries to bidders (e.g. "how much do you like this restaurant?" as opposed to "tell us how you like each restaurant in Hong Kong"). # of queries scales mildly in k and independent of m, under natural assumptions.

Example 2: Mechanism Design meets Bayesnets and MRFs

ample Based MD: n bidders, quasi-linear utilities, independent types drawn from $F = F_1 \times \cdots \times F_n$, for all i, F_i is over $\mathbb{R}^m$, and we are given ***sample access*** to F_i.

arge body of literature: [Elkind'07, Cole-Roughgarden'14, Mohri-Medina'14, Huang et al'14, Morgenstern-oughgarden'15, Devanur et al'16, Roughgarden-Schrijvers'16, Gonczarowski-Nisan'16, Goldner-Karlin'16, yrgkanis'17, Cai-Daskalakis'17, Gonczarowski-Weinberg'18, Balcan et al. '18, Guo et al. '19…]

- Many considers $m = 1$
- General m, either requires item-independence or only learn the optimal mechanism in some specific class.

Dughmi et al'14]: if distributions F_i are **arbitrarily dependent** over $\mathbb{R}^m$ then exponentially many samples in ecessary, even to attain constant-factor approximation to optimal revenue

- holds for the simple case of one unit demand bidder, m items

he instance in **[Dughmi et al'14]** requires strong dependence.

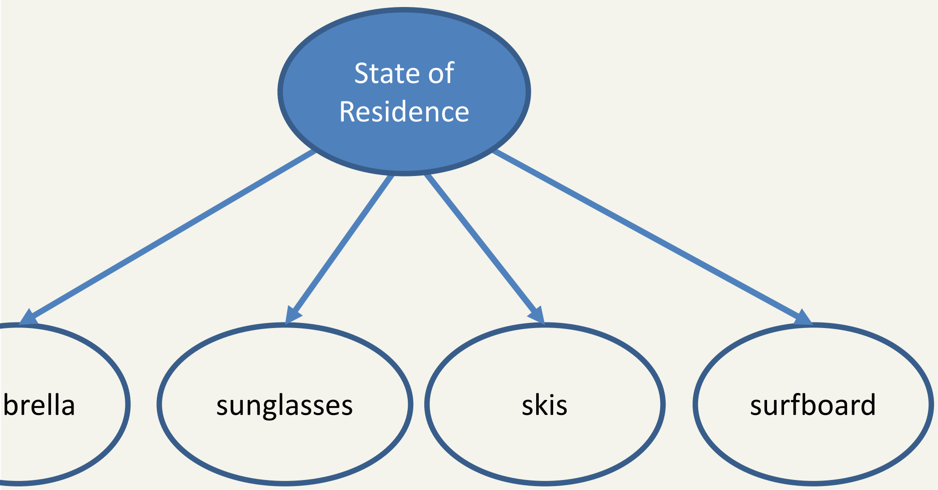
- Improve sample complexity that degrades gracefully with the degree of dependence?

Ve use two most prominent graphical models to capture dependence: **Bayesian Networks (Bayesnets)** and **M andom Fields (MRFs) .**

- Note that they are **fully general** if the graphs on which they are defined are **sufficiently dense**.
- Natural parameters of these models: **maximum size of hyperedges** in an MRF and **largest indegree** in a Bayesnet.
- Allow **latent variables**, i.e. unobserved variables in the distribution.

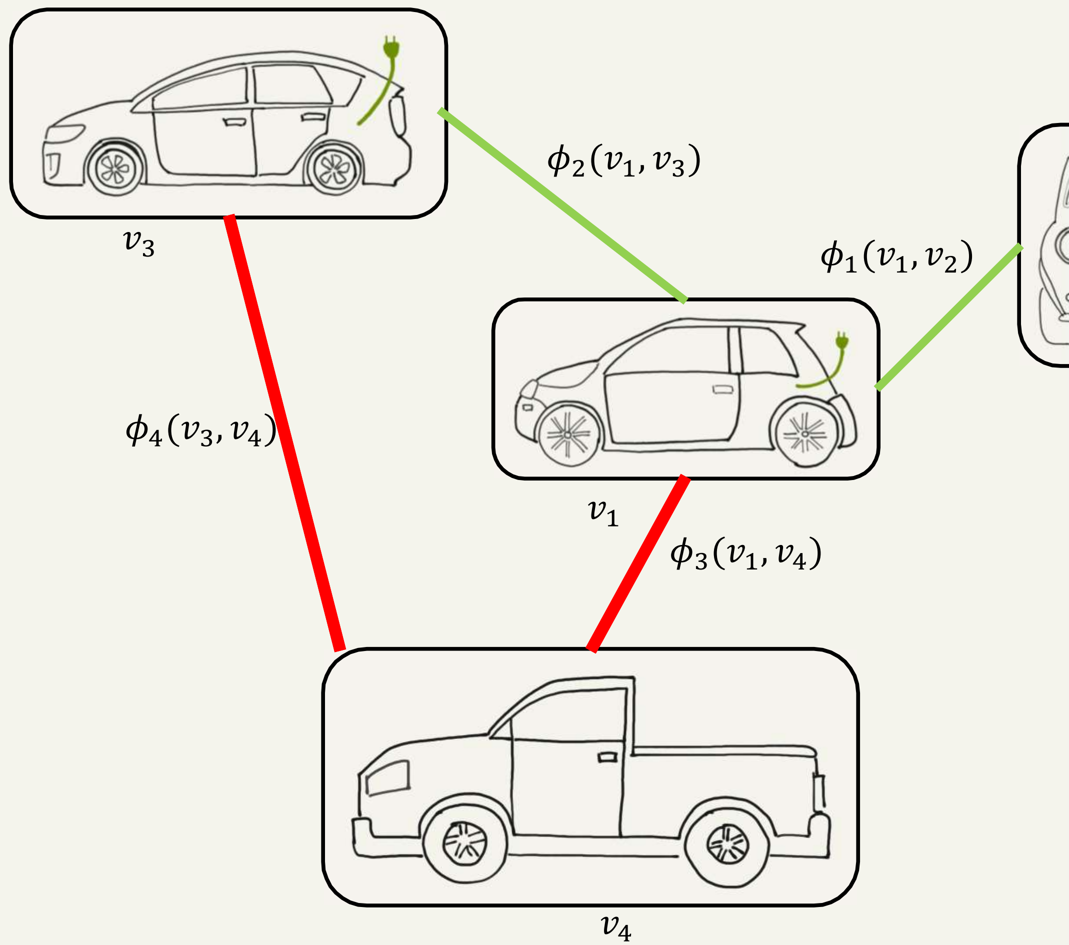
Example 2: Mechanism Design meets Bayesnets and MRFs (cont'd)

Bayesnet: a directed acyclic graph



MRF: an undirected graph

$\Pr(v_1, v_2, v_3, v_4) \propto \phi_1(v_1, v_2)\phi_2(v_1, v_3)\phi_3(v_1, v_4)\phi_4(v$



Example 2: Mechanism Design meets Bayesnets and MRFs (cont'd)

- 1+1=3 approach:
  - Step 1 (ML): learn MRF/Bayesnet $\hat{F}$ approximating true F in Prokhorov
  - Step 2 (MD): design good mechanism $\widehat{\mathcal{M}}$ for model $\hat{F}$
  - Step 3 (Robust MD): massage $\widehat{\mathcal{M}}$ into a good mechanism $\mathcal{M}_R$ for F

- Sample complexity for learning an ϵ-optimal and η-BIC mechanism:

Setting	Sample Complexity	Prior Result
Product Measure	$\text{poly}\left(n, m, \frac{1}{\epsilon}, \frac{1}{\eta}\right)$	[Gonczarowski-Weinberg '18]
MRF (d=max clique size)	$\text{poly}\left(n, m^d, \|\Sigma\|^d, \frac{1}{\epsilon}, \frac{1}{\eta}\right)$	unknown
Bayesnet (d=max indegree)	$\text{poly}\left(n, d, m, \|\Sigma\|^d, \frac{1}{\epsilon}, \frac{1}{\eta}\right)$	

n=#bidders, m=#items, $\Sigma =$ effective value range

Exponential dependence on d is unavoidable as $d = \Omega(m)$ allows full dependence [Dughmi et al'14]

Conclusion

- Main Result: Max-Min Mechanism Design Robustness Under Prokhorov in multi-dimensional settings.

- A new modular approach to MD
  - Learn model $\hat{F}$ to within some distance; Prokhorov is good.
  - Find good mechanism $\widehat{\mathcal{M}}$ for $\hat{F}$.
  - Massage $\widehat{\mathcal{M}}$ to $\mathcal{M}_R$ that is robust to the model misspecification.

- I think we are at a turning point for MD + ML
  - we have a modular framework that allows disentangling the two.
  - lots of opportunities in ML meets MD space.

Thank you!