Mixed Integer Programming Approaches for Experimental Design

Juan Pablo Vielma

Massachusetts Institute of Technology

Joint work with

Chris Coey, Miles Lubin and Denis Saure

Workshop in Management Science.

Puerto Varas, Chile, 2017.

Experimental Design for Preference Surveys

Feature	SX530	RX100
Zoom	50x	3.6x
Prize	$249.99	$399.99
Weight	15.68 ounces	7.5 ounces
Prefer	☑	☐

Feature	TG-4	G9
Waterproof	Yes	No
Prize	$249.99	$399.99
Weight	7.36 lb	7.5 lb
Prefer	☐	☑

Feature	TG-4	Galaxy 2
Waterproof	Yes	No
Prize	$249.99	$399.99
Viewfinder	Electronic	Optical
Prefer	☑	☐

Parametric Preference Model

↕

Estimate Preference Parameter

Experimental Design for Preference Surveys

Feature	Chewbacca	BB-8
Wookiee	Yes	No
Droid	No	Yes
Blaster	Yes	No
I would buy toy	☑	☐
Product Profile	x^1	x^2

$$\begin{pmatrix}0\\1\\0\end{pmatrix} = x^2$$

- Even evaluating “quality” of survey design may be expensive:
  – Can state-of-the-art MIP help?

50+ Years of MIP = Significant Solver Speedups

- Machine-independent algorithmic improvements (on standard and solver benchmark instances):
  - **CPLEX** v1.2 (1991) – v11 (2007): 29,000x speedup
  - **Gurobi** v1 (2009) – v6.5 (2015): 48.7x speedup
  - Commercial, but free for academic use (Also Xpress)
- (Reasonably) effective free / open source solvers:
  - GLPK, **COIN-OR (CBC)** and SCIP (only for non-commercial)
- Accessible, fast and versatile 21st century tools:
  - julia -based JuMP modelling language
- Mature and evolving effectiveness:
  - **Linear MIP**, second order cone MIP (**MI-SOCP**) and convex/conic nonlinear MIP (**MI-SDP**, **MI-SDP+EXP**)

Case Study 1:

Quick Linear Regression
During Christmas in Viña del Mar

Experimental Design for Linear Regression

Model: $y^i = \beta \cdot z^i + \epsilon_i, \quad \epsilon_i \sim N(0,1)$

Questions:

$$Z = \left[z^1 | \dots | z^q\right]^T \in \mathbb{R}^{q \times n}$$

Answers:

$$Y = \left[y_1 | \dots | y_q\right]^T$$

(One) design goal = Min. "variance" of estimator of $\beta \in \mathbb{R}^n$

(One) Version of variance for OLS = D-efficiency :

$$\max_{Z \in \mathcal{Z}} \left(\det\left(Z^T Z\right)\right)^{1/q}$$

Example $\mathcal{Z}$ = Product profiles

$$x^{1,i}, x^{2,i} \in \{0,1\}^n$$

$$z^i = x^{1,i} - x^{2,i} \in \{-1,0,1\}^n$$

MIP = flexible $\mathcal{Z}$. e.g. partial profiles : $\left\|x^{1,i} - x^{2,i}\right\|_1 \leq m$

MIP Formulations Approaches

- Traditional MI-SDP (SDP representation of $\det^{1/q}$):

$$\{z^{i_j}\}_{j=1}^{k} \to \max_{w} \left\{ \left(\det \left(\sum_{j=1}^{k} w_j z^{i_j} \cdot {z^{i_j}}^T \right) \right)^{1/q} : \begin{array}{r} \sum_{j=1}^{k} w_j = q \\ w \in \mathbb{Z}_+^k \end{array} \right\}$$

- MI-SOCP reformulation (Sagnol and Harman, '15)
- MI-SDP + linearization of products of binaries:

$$\max_{z^i, x^{1,i}, x^{2,i}} \left\{ \left(\det \left(\sum_{i=1}^{q} z^i \cdot {z^i}^T \right) \right)^{1/q} : \begin{array}{r} x^{1,i} - x^{2,i} = z^i \\ x^1, x^2 \in \{0,1\}^n \end{array} \right\}$$

- MI-SDP+EXP :

$$\left(\det \left(\sum_{i=1}^{q} z^i \cdot {z^i}^T \right) \right)^{1/q} \to \log \det \left(\sum_{i=1}^{q} z^i \cdot {z^i}^T \right)$$

Solvers for Mixed Integer Conic Programming

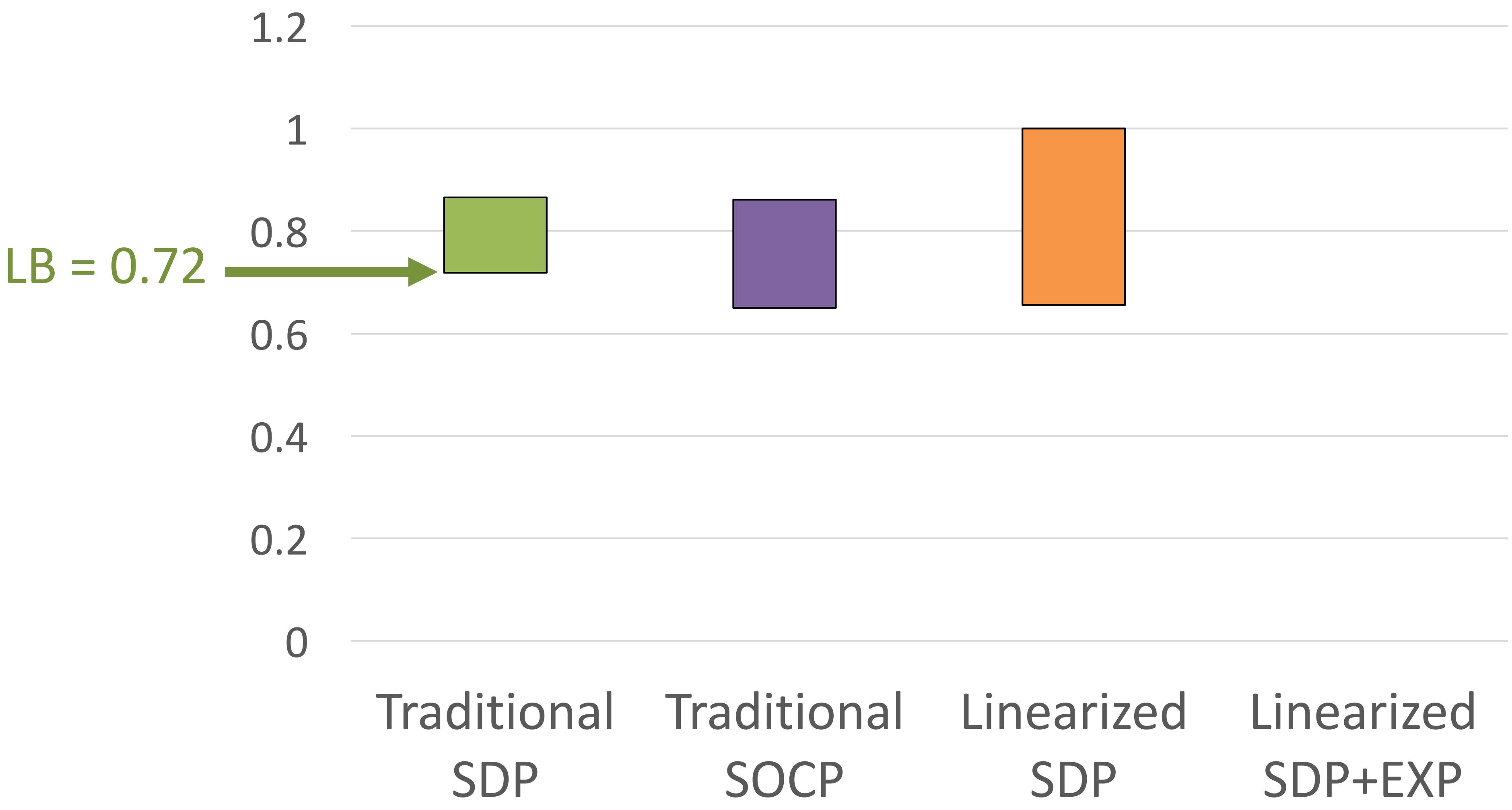
- **MI-SOCP** : relatively mature + active development:
  - V., Dunning, Huchette and Lubin '16: **Extended formulations**. Adopted by Gurobi 6.5 (**4x speedup**). Also adopted in CPLEX 12.6.3 and Xpress 8.0.
- **MI-SDP** : only basic algorithms until:
  - **Pajarito**: Lubin, Yamangil, Bent and V. '16 and Coey, Lubin and V. '17. Uses generic **linear-MIP** and **conic solver**.

  - SCIP-SDP: Gally, Pfetsch and Ulbrich '16 (harder to install).
- **MI-SDP+EXP** : Also **Pajarito**:
  - Less stable (IPM conic solver) and needs CVX/Convex.jl

Computational Experiment

- 12 binary features and 16 questions.
- Random k = 500 for traditional.
- **SOCP** = **Gurobi** 7
- **MI-SDP** = **Pajarito** with **Gurobi** 7 and **Mosek** 8
- **MI-SDP+EXP** = **Pajarito** with **Gurobi** 7 and **SCS** 1.1.8
- Core i7-6700K CPU @ 4.00GHz, 32GB RAM (Latest iMac)

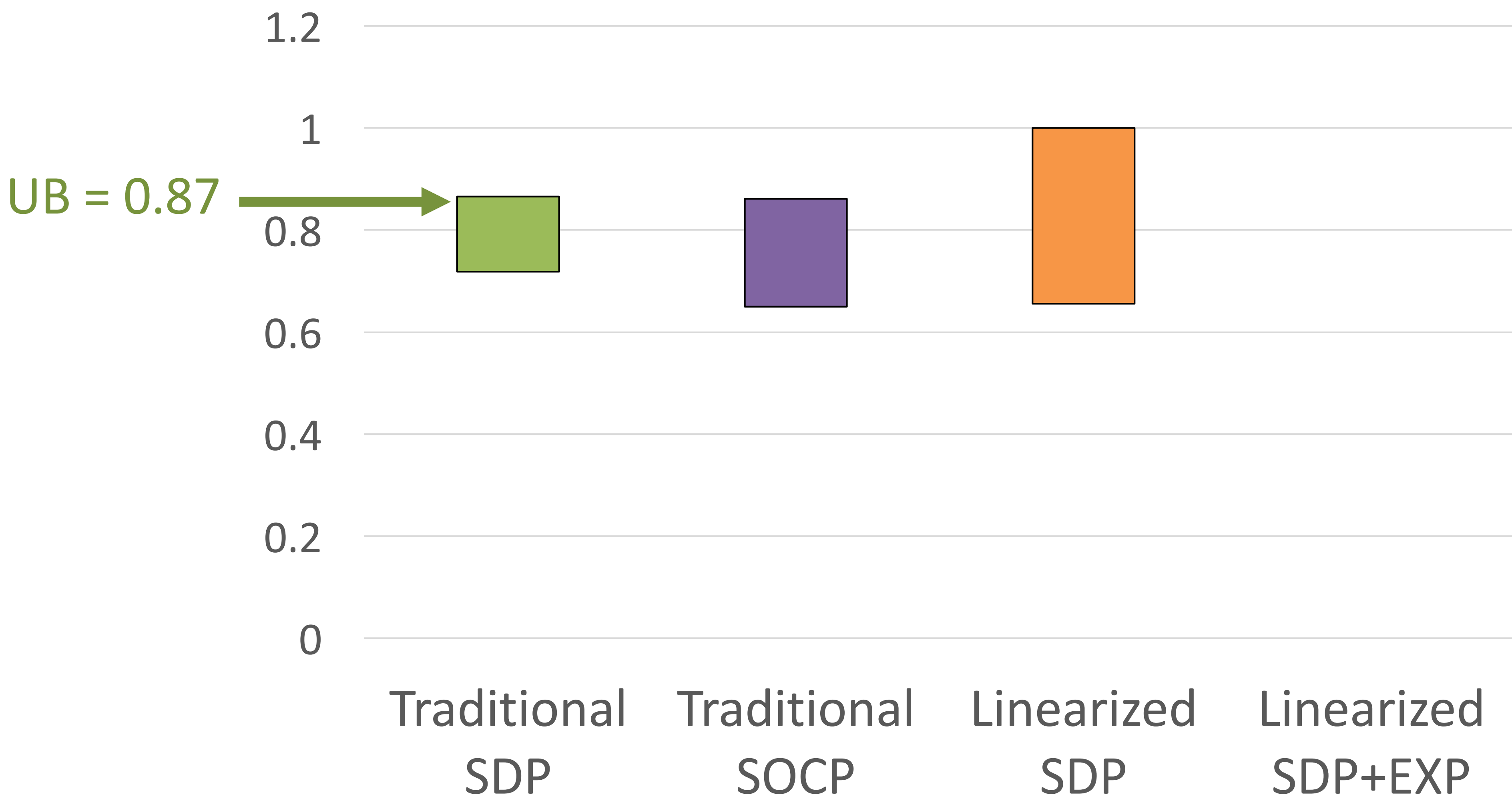
Get a cup of coffee time length = 5 min



Lower Bound = Feasible Solution.

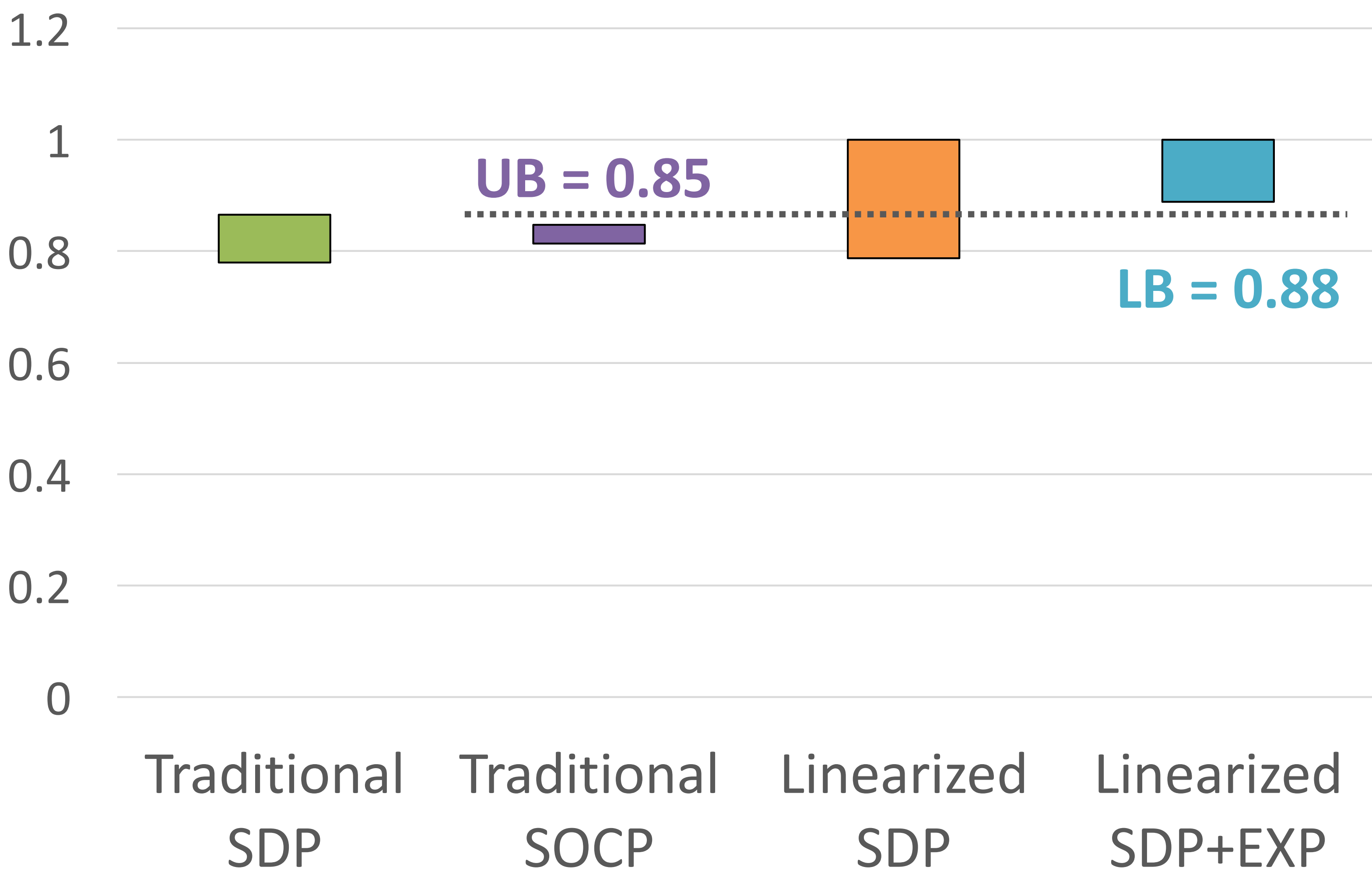
e.g. Traditional SDP yields a design with D-eff = 0.72

Get a cup of coffee time length = 5 min

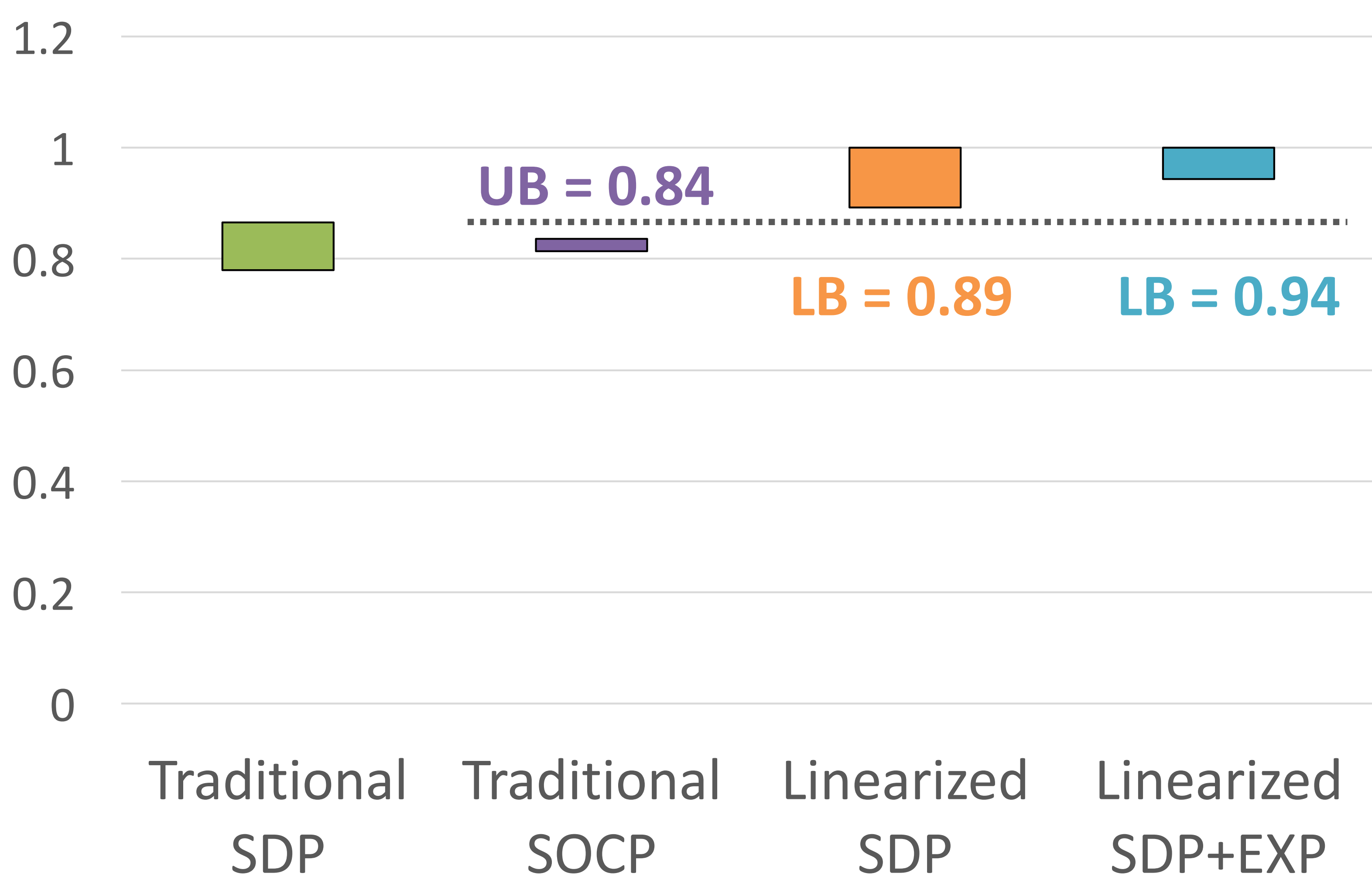


Upper Bound = Bound on best possible solution (**for model**).
e.g. Any design **for traditional SDP** has D-eff ≤ 0.87

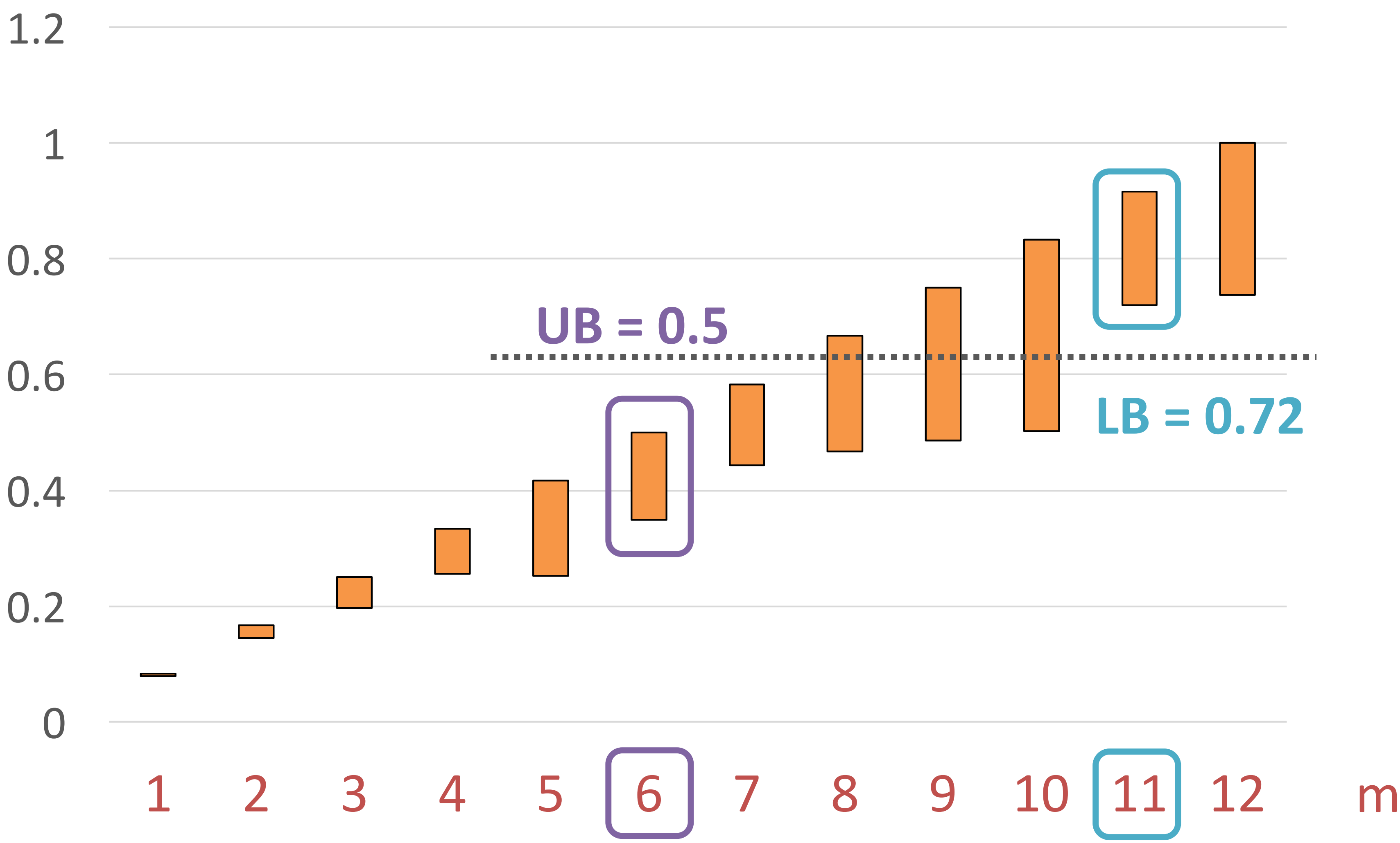
Going to lunch time length = 1 hour



Overnight time length = 16 hour

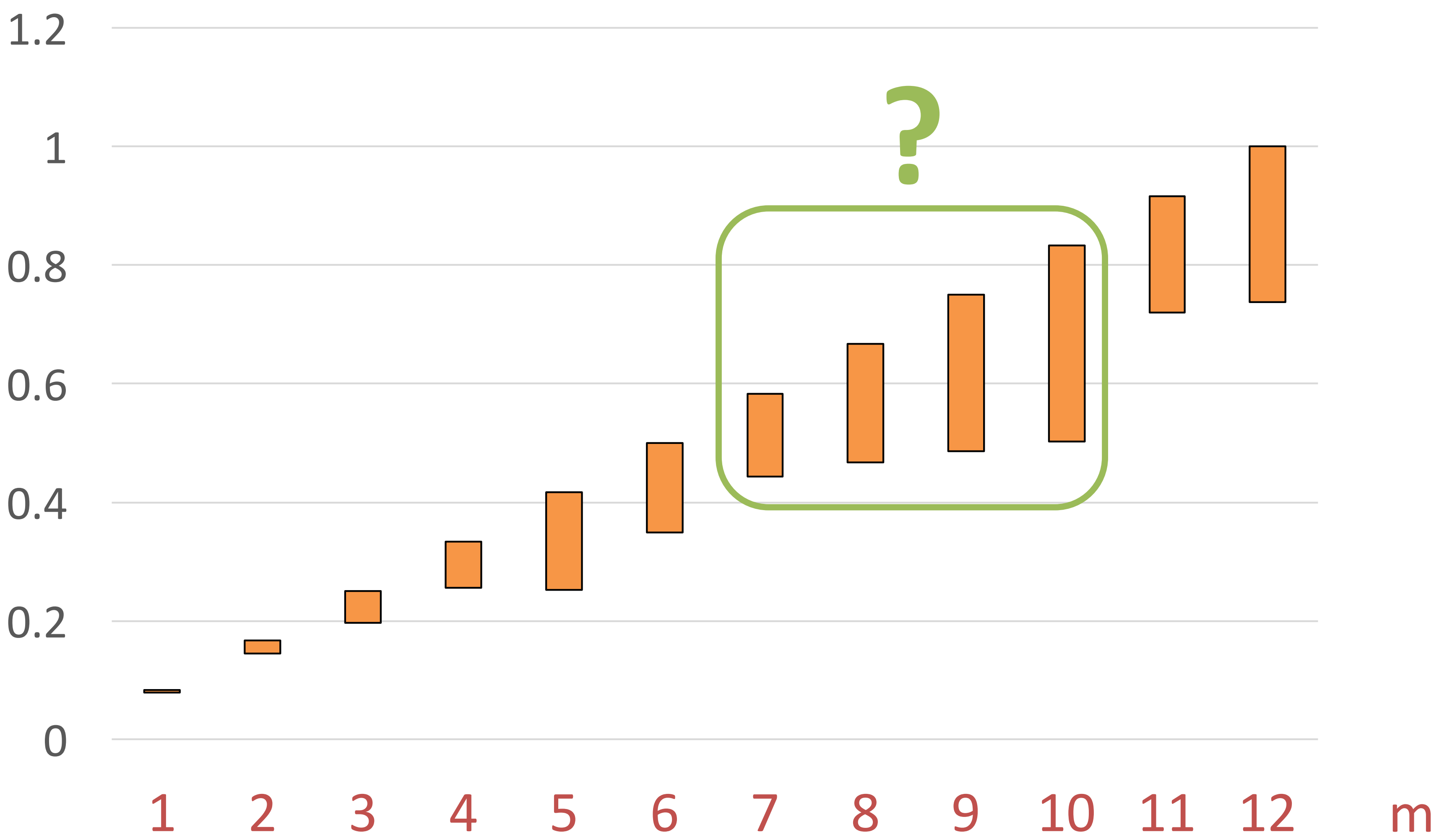


Application: Cost of Partial Profiles (L. SDP – 1 h)



Partial Profile: 2 products differ in only m features

Application: Cost of Partial Profiles (L. SDP – 1 h)



Partial Profile: 2 products differ in only m features

Case Study 2:

Choice-Based Comparisons

and

Real-Time Adaptive Logistic Regression

Choice-based Conjoint Analysis

Feature	Chewbacca	BB-8
Wookiee	Yes	No
Droid	No	Yes
Blaster	Yes	No
I would buy toy	☑	☐
Product Profile	x^1	x^2

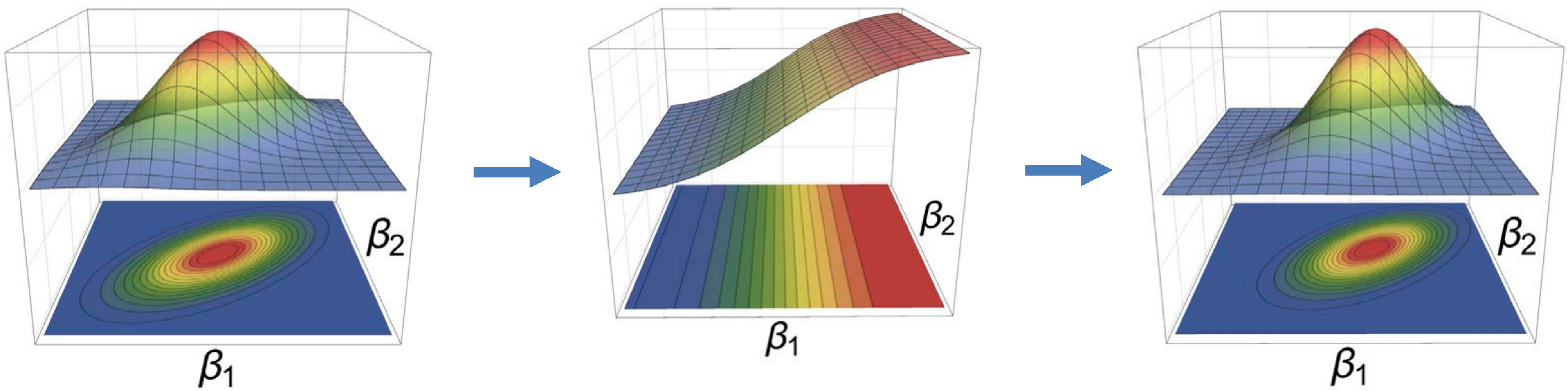
$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = x^2$$

MNL Preference Model ⟷ Logistic Regression

$$\beta \cdot x^1 \geq \beta \cdot x^2 \quad \Leftrightarrow \quad \beta \cdot z \geq 0 \quad z = x^1 - x^2$$

1-Question Bayesian Logistic Regression

Prior distribution — Answer likelihood — Posterior distribution



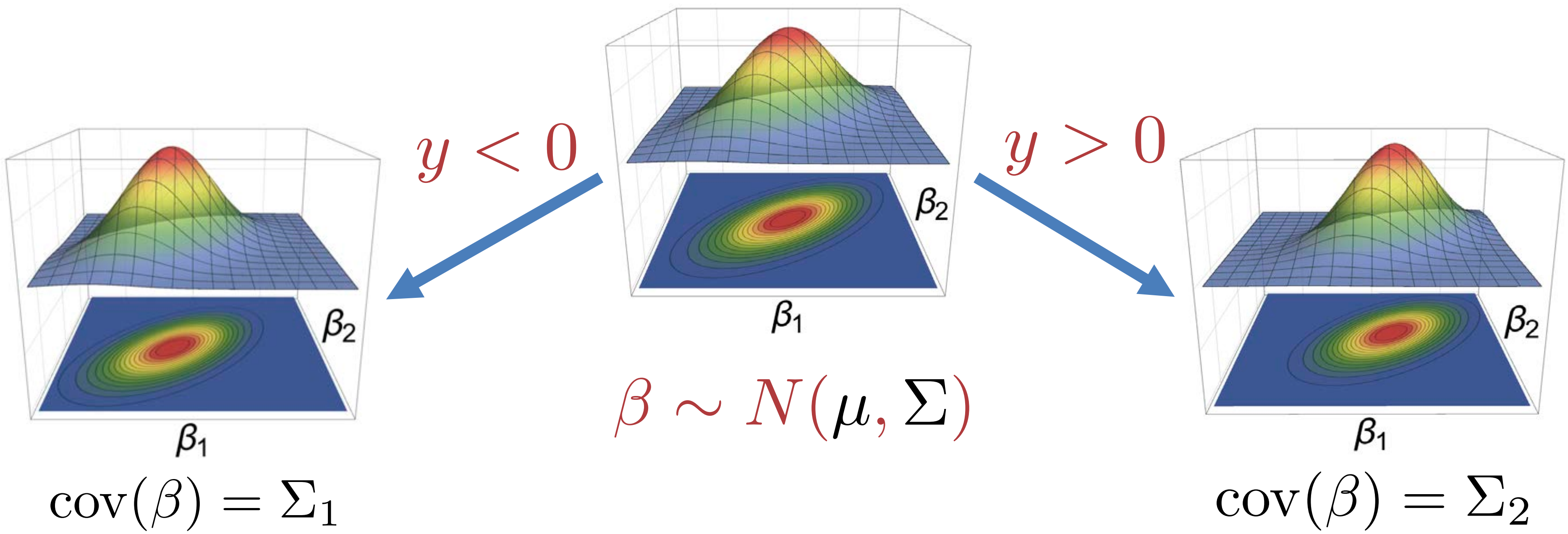
$$\beta \sim N(\mu, \Sigma) \qquad L\,(y \mid \beta, z) \qquad g\,(\beta \mid y, z\,)$$

$$g\,(\beta \mid y, z\,) \propto \phi\,(\beta\,;\; \mu, \Sigma)\, L\,(y \mid \beta, z)$$

$$L\,(y \mid \beta, z) = \left(1 + e^{-\beta \cdot z}\right)^{-1}$$

D-Efficiency and Expected Posterior Variance

$$f(z, \mu, \Sigma) := \mathbb{E}_{y,\beta}\left\{(\det \operatorname{cov}(\beta \,|\, y, z))^{1/m}\right\}$$



$$\max_{z \in \{-1,0,1\}^n} f(z, \mu, \Sigma)$$

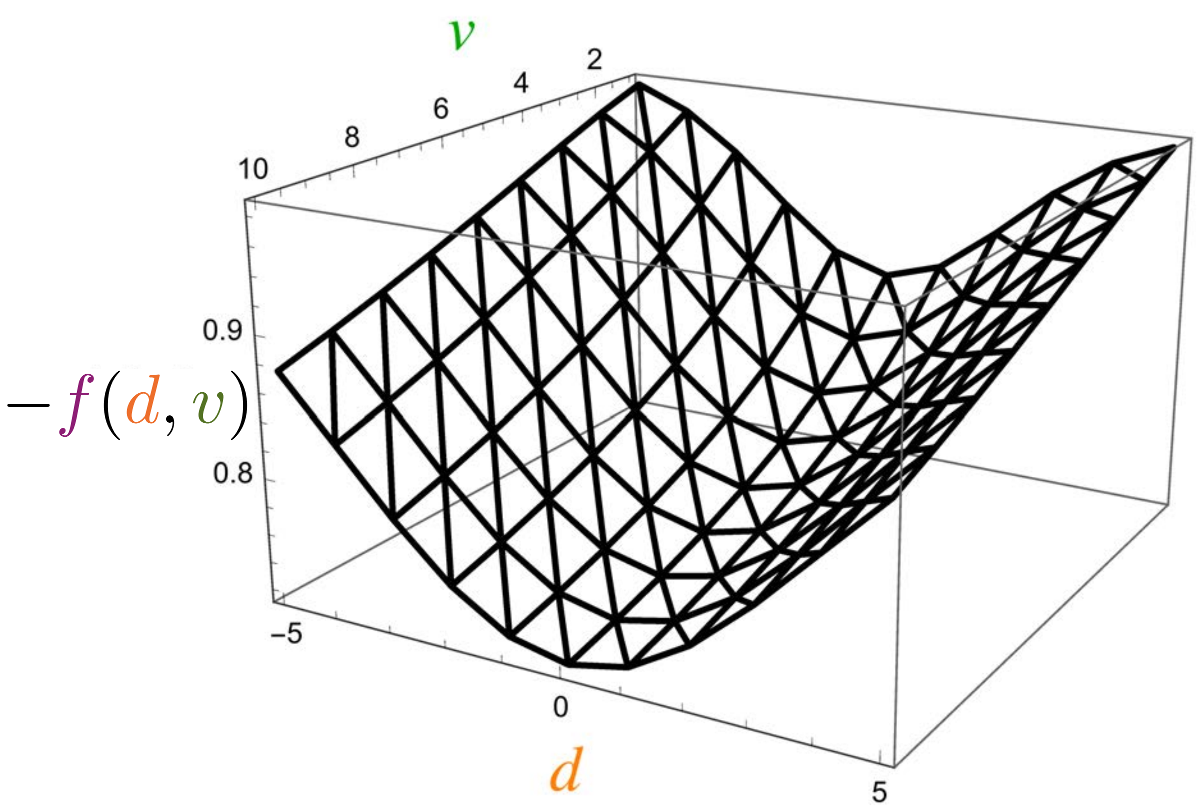
- $f(z, \mu, \Sigma)$ is hard to evaluate, non-convex and n large

Reformulation from V. and Saure '16

- D-efficiency $f(z)$ = Non-convex function $f(d, v)$ of

$$\text{mean:} \quad d := \mu \cdot z$$

$$\text{variance:} \quad v := z' \cdot \sum \cdot z$$



Can evaluate $f(d, v)$ with 1-dim integral 🙂

Piecewise Linear Interpolation

Linear MIP formulation

Aligns with selection criteria from Toubia, Hauser, and Simester '04: minimize mean and maximize variance

MIP-based Moment-Matching Approx. Bayes


Feature	SX530	RX100
Zoom	50x	3.6x
Prize	$249.99	$399.99
Weight	15.68 ounces	7.5 ounces
Prefer	☑	☐


Feature	TG-4	G9
Waterproof	Yes	No
Prize	$249.99	$399.99
Weight	7.36 lb	7.5 lb
Prefer	☐	☑


Feature	TG-4	Galaxy 2
Waterproof	Yes	No
Prize	$249.99	$399.99
Viewfinder	Electronic	Optical
Prefer	☑	☐

$$\mathbb{E}\left(\beta \,|\, Y, X^1, X^2\right)$$

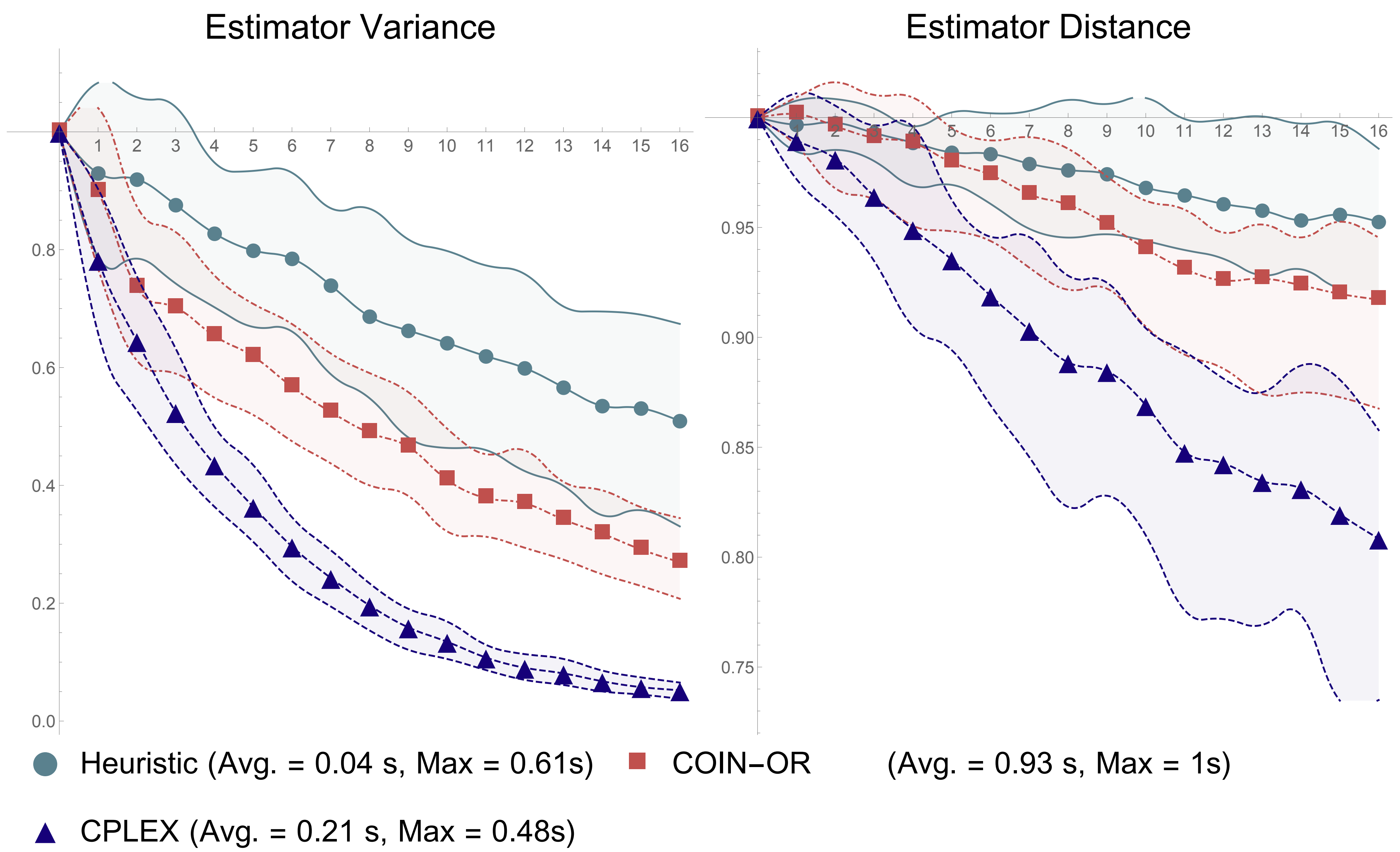
$$\operatorname{cov}\left(\beta \,|\, Y, X^1, X^2\right)$$

- After each answer compute posterior mean and covariance and replace with corresponding Gaussian

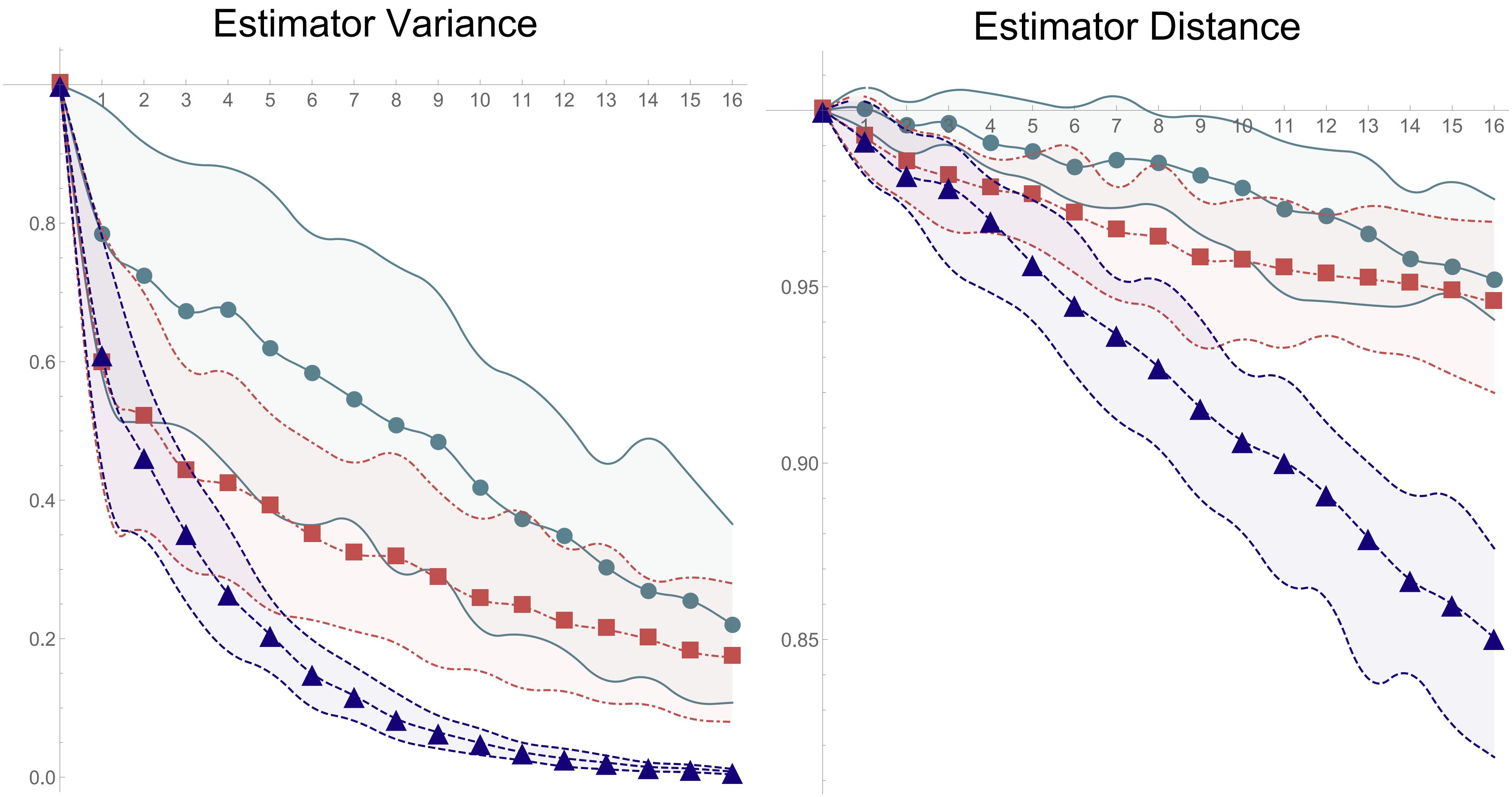
Computational Experiments

- 16 questions, 2 options, 12 and 24 features
- Simulate MNL responses with known β^*
- Question Selection
  - Linear MIP-based using **CPLEX** and open source **COIN-OR**
  - Knapsack-based geometric **Heuristic** by Toubia et al.
- Time limits of 1 s and 10 s
- Metrics:
  - Estimator variance = $\left(\det \operatorname{cov}\left(\beta \,|\, Y, X^1, X^2\right)\right)^{1/2}$
  - Estimator distance = $\left\|\mathbb{E}\left(\beta \,|\, Y, X^1, X^2\right) - \beta^*\right\|_2$
  - Computed for true posterior with MCMC
- Slightly slower computer (~ '12 iMac)

Results for 12 Features, 1 s time limit



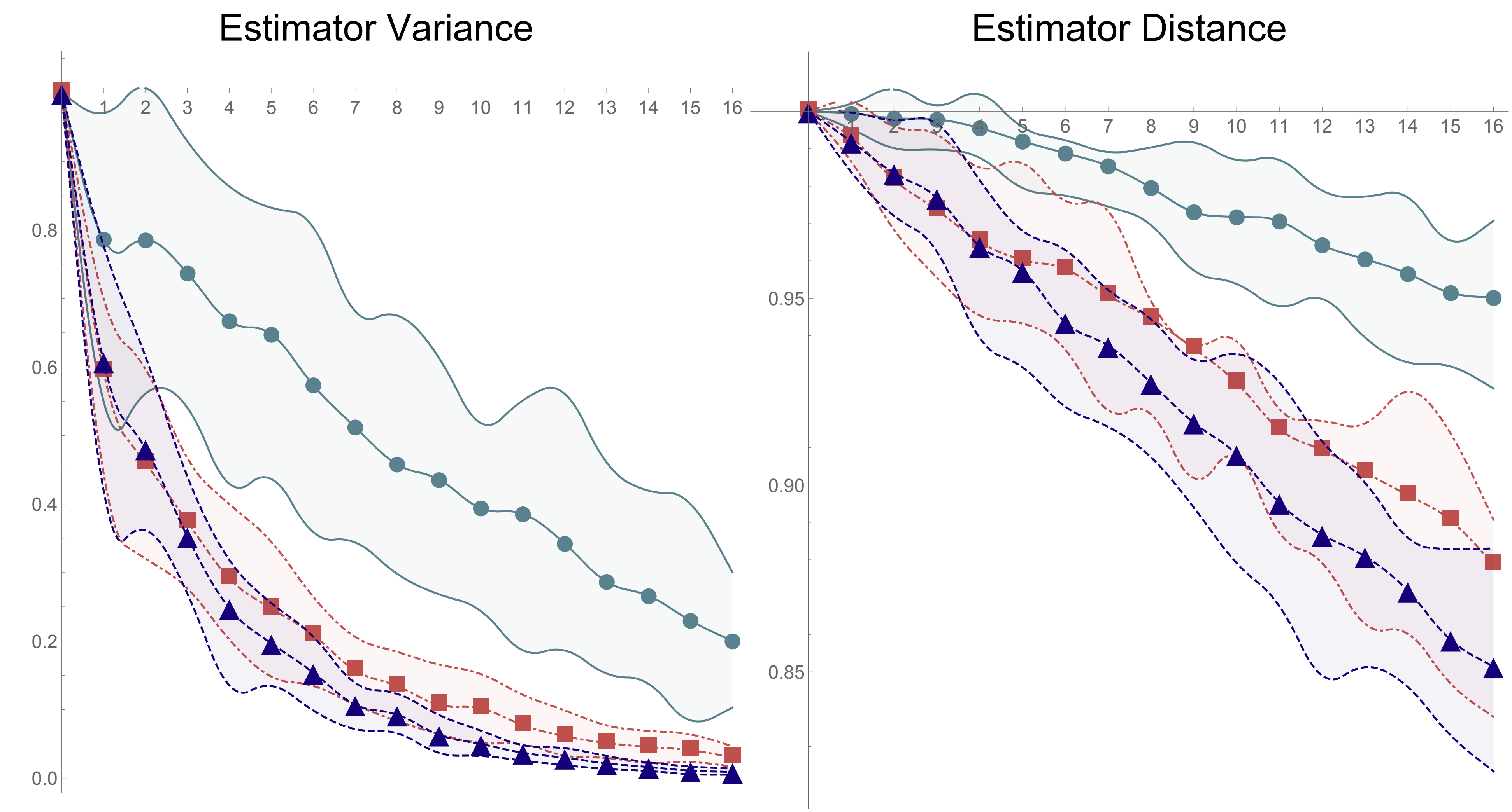
Does it Scale? Results for 24 features



● Heuristic (Avg. = 0.19 s, Max = 3s) ■ CPLEX 1s (Avg. = 1 s, Max = 1s)

▲ CPLEX 10s (Avg. = 7.7 s, Max = 10s)

Some improvements for 24 features



● Heuristic (Avg. = 0.19 s, Max = 3s) ■ CPLEX 1s (Avg. = 1 s, Max = 1s)

▲ CPLEX 10s (Avg. = 7.7 s, Max = 10s)

Summary and Extensions

- MIP for experimental design
  - Effective even for non-linear and near-real time
  - Appropriate domain expertise can be crucial for MIP'ing
  - Commercial solvers best, but free solvers reasonable
  - Integration into complex systems easy with JuMP
  - Some scalability : get the most out of "small" data
- Multi-Question Bayesian Logistic Regression: $Z = \{z^i\}_{i=1}^q$
  - For many variants and approximations of "variance"

$$f(Z, \mu, \Sigma) = f\left(\{\mu \cdot z^i\}_{i=1}^q, \left\{{z^i}^T \Sigma z^j\right\}_{i,j=1}^q\right)$$

- What kind of designs do **you** want to build? (future benchmark instances)