Mixed Integer Programming Approaches for Experimental Design

Juan Pablo Vielma

Massachusetts Institute of Technology

Joint work with Denis Saure

DRO brown bag lunch seminars, Columbia Business School
New York, NY, October, 2016.

Motivation: (Custom) Product Recommendations

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	☑	☐

We recommend:

Motivation: (Custom) Product Recommendations

Feature	SX530	RX100
Zoom	50x	3.6x
Prize	$249.99	$399.99
Weight		
Prefer		

Feature	TG-4	G9
Waterproof	Yes	No
Prize	$249.99	$399.99
		7.5 lb
		☑

- Toubia, Hauser and Dahan (2003)
- Toubia, Hauser, and Simester (2004)

nd:

Feature		
Waterproof	Yes	No
Prize	$249.99	$399.99
Viewfinder	Electronic	Optical
Prefer	☑	☐

“Towards” Optimal Product Recommendation

- Find enough information about preferences to recommend

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	☐	☑

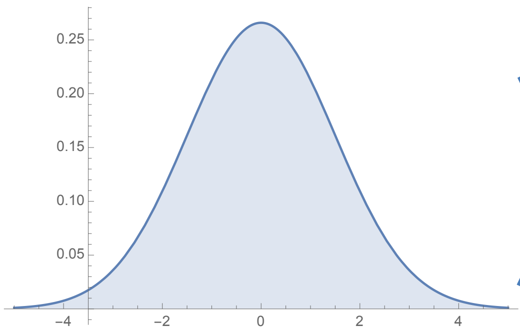
Feature	TG-4	Galaxy 2
Waterproof	Yes	No
Prize	$249.99	$399.99
Viewfinder	Electronic	Optical
Prefer	☑	☐

β

- How do I pick the **next (1st) question** to obtain the largest reduction of uncertainty or “variance” on preferences
- Compensatory model estimation **(part-worths)**, not just assortment

Next Question To Reduce "Variance": Bayesian

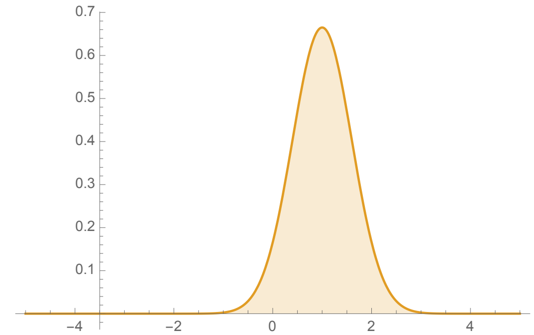
Prior Distribution of β



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

Bayesian Update

Posterior Distribution

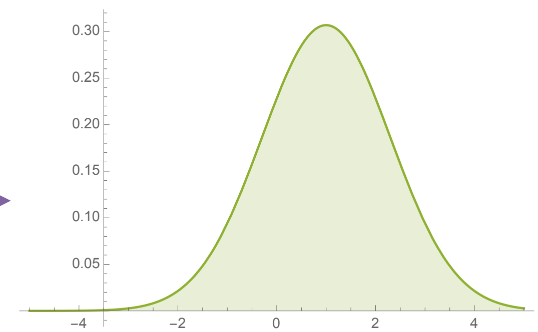


MCMC

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

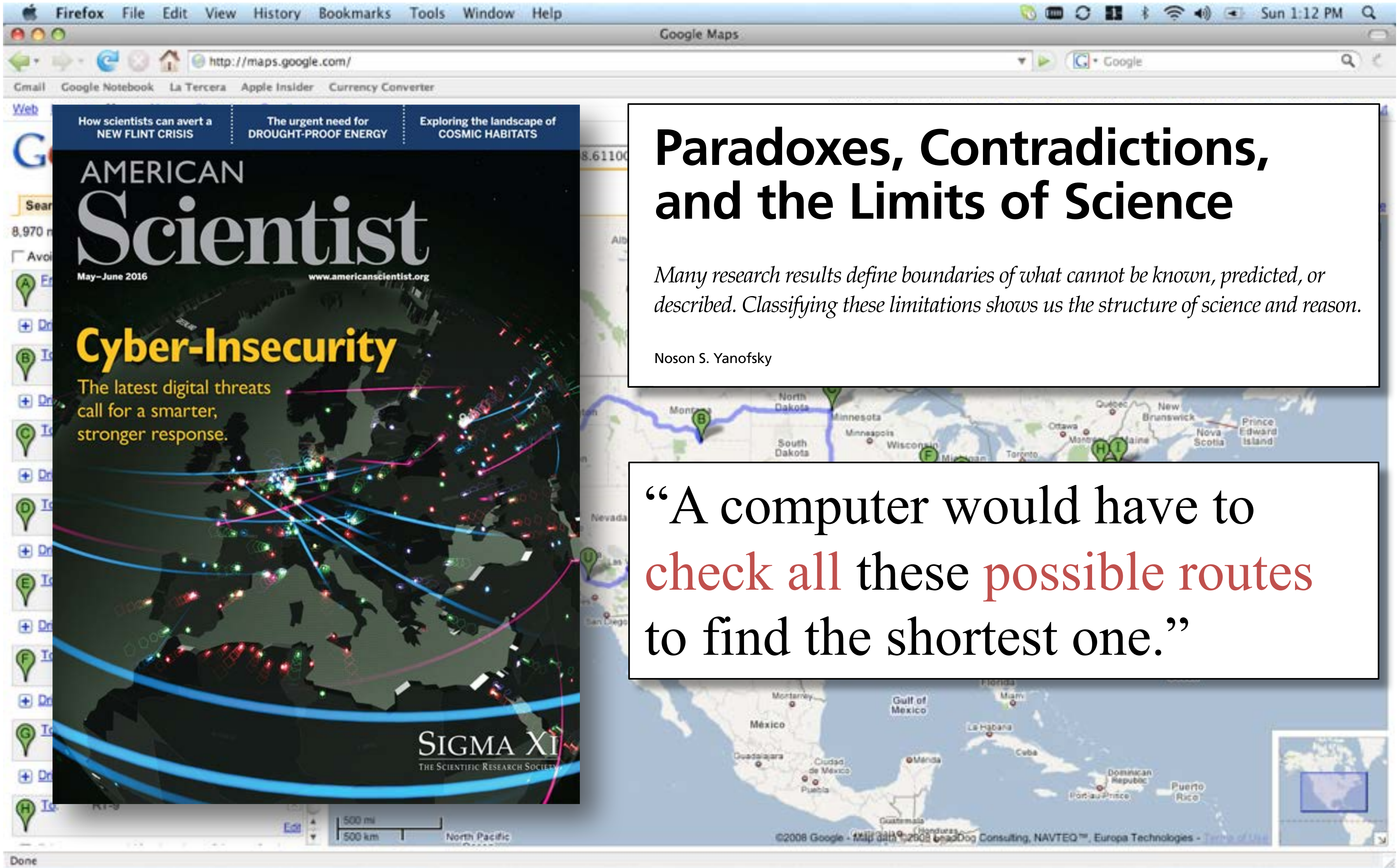
Bayesian Update

Posterior Distribution



- Black-box objective: Question Selection = Enumeration ☹
- Question selection by Mixed Integer Programming (MIP)

Traveling Salesman Problem (TSP): Visit Cities Fast



Paradoxes, Contradictions, and the Limits of Science

Many research results define boundaries of what cannot be known, predicted, or described. Classifying these limitations shows us the structure of science and reason.

Noson S. Yanofsky

“A computer would have to check all these possible routes to find the shortest one.”

MIP = Avoid Enumeration

- Number of tours for 49 cities $= 48!/2 \approx 10^{60}$
- Fastest supercomputer $\approx 10^{17}$flops
- Assuming one floating point operation per tour:
  $> 10^{35}$ years $\approx 10^{25}$ times the age of the universe!
- How long does it take on an iphone?
  - Less than a second!
  - 4 iterations of **cutting plane** method!
  - Dantzig, Fulkerson and Johnson 1954 did it by hand!
  - For more info see tutorial in ConcordeTSP app
  - **Cutting planes** are the key for effectively solving (even NP-hard) MIP problems in practice.

50+ Years of MIP = Significant Solver Speedups

- Algorithmic Improvements (Machine Independent):
  - **CPLEX** v1.2 (1991) – v11 (2007): 29,000x speedup
  - Gurobi v1 (2009) – v6.5 (2015): 48.7x speedup
  - Commercial, but free for academic use
- (Reasonably) effective free / open source solvers:
  - GLPK, **COIN-OR (CBC)** and SCIP (only for non-commercial)
- Easy to use, fast and versatile modeling languages
  - Julia based JuMP modelling language
- Linear MIP solvers very mature and effective:
  - Convex nonlinear MIP getting there (even MI-SDP!), quadratic nearly there

Choice-based Conjoint Analysis (CBCA)

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

- Utilities for 2 products, n features (e.g. n = 12)

$$U_1 = \beta \cdot x^1 + \epsilon_1 = \sum_{i=1}^{n} \beta_i x_i^1 + \epsilon_1$$

$$U_2 = \beta \cdot x^2 + \epsilon_2 = \sum_{i=1}^{n} \beta_i x_i^2 + \epsilon_2$$

part-worths

product profile

noise (gumbel)

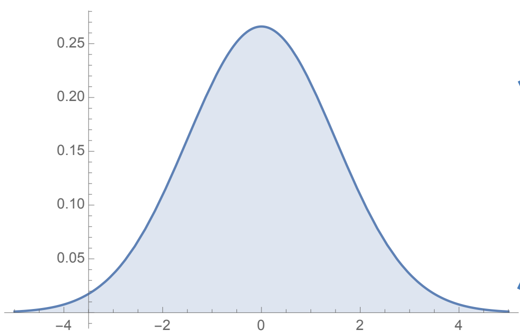
- Utility maximizing customer: $x^1 \succeq x^2 \Leftrightarrow U_1 \text{``}\geq\text{''} U_2$

- Noise can result in response error:

$$L\left(\beta \,|\, x^1 \succeq x^2\right) = \mathbb{P}\left(x^1 \succeq x^2 \mid \beta\right) = \frac{e^{\beta \cdot x^1}}{e^{\beta \cdot x^1} + e^{\beta \cdot x^2}}$$

Next Question To Reduce "Variance": Bayesian

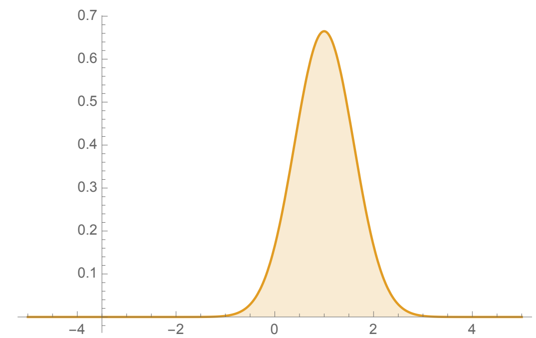
Prior Distribution of β



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

Bayesian Update

Posterior Distribution

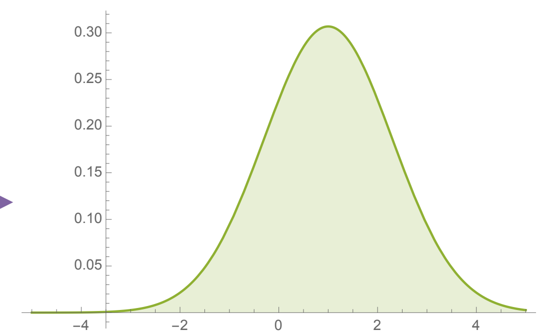


MCMC

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

Bayesian Update

Posterior Distribution



MNL Preference Model ⟷ Logistic Regression

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

"Linear" Experimental Design

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

Unknown $\beta \in \mathbb{R}^n$

$$\left\{ \beta \cdot z^i \right\}_{i=1}^{q}$$

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

Questions:

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

Answers:

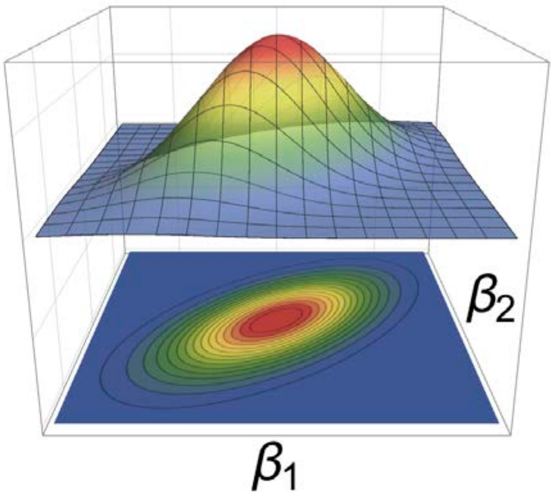
$$Y = \left[y_1 | \ldots | y_q \right]^T$$

Model: $\mathbb{P}\left(Y \mid \beta, Z\right) = L\left(Y \mid \beta, Z\right) = \prod_{i=1}^{q} h\left(y^i, \beta \cdot z^i\right)$

Objective: Choose Z to learn β "fast"

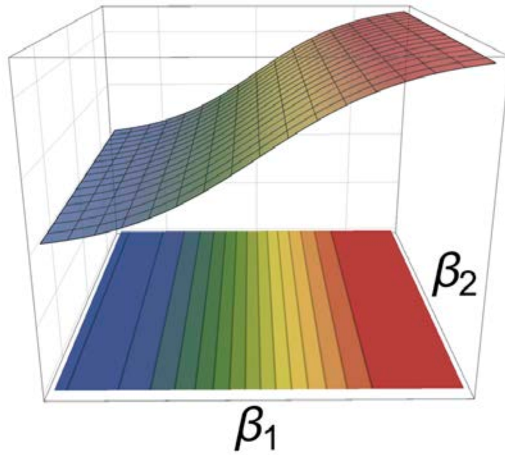
Bayesian Framework

Prior distribution



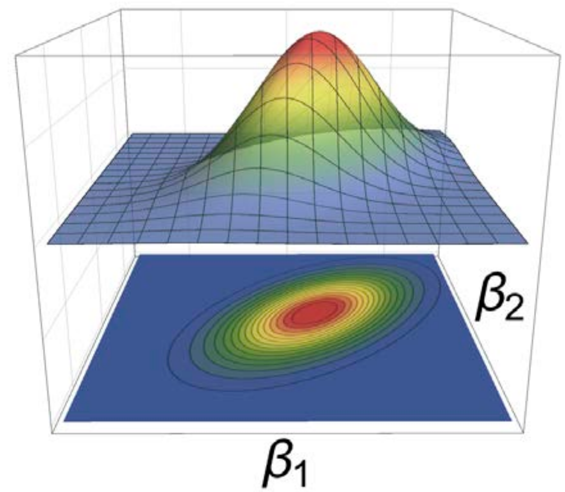
$$\beta \sim N(\mu, \Sigma)$$

Answer likelihood



$$L\,(Y \mid \beta, Z)$$

Posterior distribution



$$g\,(\beta \mid Y, Z\ )$$

$$g\,(\beta \mid Y, Z\ ) = \frac{\phi\,(\beta\ ;\ \mu, \Sigma)\, L\,(Y \mid \beta, Z)}{\int_{\mathbb{R}} \phi\,(\beta\ ;\ \mu, \Sigma)\, L\,(Y \mid \beta, Z)\, d\beta}$$

“fast” $=$ minimize posterior “variance”

Goal = Minimize Expected Posterior Variance $f\left(Z\right)$

$$f\left(Z\right) = \mathbb{E}_Y \left\{ \left(\det \operatorname{cov}\left(\beta \,|\, Y, Z\right)\right)^{1/m} \right\}$$

Possible solution approaches:

$$g\left(\beta \mid Y, Z\ \right) \propto \phi\left(\beta\ ;\ \mu, \Sigma\right) L\left(Y \mid \beta, Z\right)$$

$$I\left(\beta \,|\, Y, Z\right) := -\frac{\partial^2}{\partial\beta\partial\beta} \ln g\left(\beta \mid Y, Z\ \right) \propto \Sigma^{-1} - \frac{\partial^2}{\partial\beta\partial\beta} \ln L\left(Y \mid \beta, Z\right)$$

$$\operatorname{cov}\left(\beta \,|\, Y, Z\right) \approx I\left(\hat{\beta} \,|\, Y, Z\right)^{-1}, \quad \mathbb{E}_{\beta \sim N(\mu,\sigma)} \left\{ I\left(\beta \,|\, Y, Z\right)^{-1} \right\}$$

$$\max_Z \mathbb{E}_Y \left\{ \left(\det I\left(\hat{\beta} \,|\, Y, Z\right)\right)^{1/m} \right\} \quad ?$$

A Really Good Case = Linear Regression

$$f(Z) = \mathbb{E}_Y \left\{ (\det \operatorname{cov}(\beta \mid Y, Z))^{1/m} \right\}$$

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

$$g(\beta \mid Y, Z) = \phi(\beta \,;\, \mu', \Sigma')$$

$$\Sigma' = \operatorname{var}(\beta \mid Y, Z) = \left(Z^T Z + \Sigma^{-1}\right)^{-1}$$

$$\min_Z f(Z) = \max_Z \left(\det\left(Z^T Z + \Sigma^{-1}\right)\right)^{1/m}$$

Z discrete $\longrightarrow$ MISDP or MISOCP for m = n

A Relatively Good Case = Few Questions

$$f(Z) = \mathbb{E}_Y \left\{ \left(\det \operatorname{cov}\left(\beta \,|\, Y, Z\right) \right)^{1/m} \right\}$$

$$Z = \left\{ z^i \right\}_{i=1}^{q} \subseteq \mathbb{R}^n, \qquad q \ll n$$

$$\mathbb{E}\left(\beta \,|\, Y, Z\right) = m \left(Y, \left\{ \mu \cdot z^i \right\}_{i=1}^{q}, \left\{ {z^i}^T \Sigma z^j \right\}_{i,j=1}^{q} \right)$$

$$\operatorname{cov}\left(\beta \,|\, Y, Z\right) = M \left(Y, \left\{ \mu \cdot z^i \right\}_{i=1}^{q}, \left\{ {z^i}^T \Sigma z^j \right\}_{i,j=1}^{q} \right)$$

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

A Relatively Good Case = Few Questions

$$f(Z) = \mathbb{E}_Y\left\{\left(\det \operatorname{cov}\left(\beta \,|\, Y, Z\right)\right)^{1/m}\right\}$$

$Z =$

$\mathbb{E}(\beta$

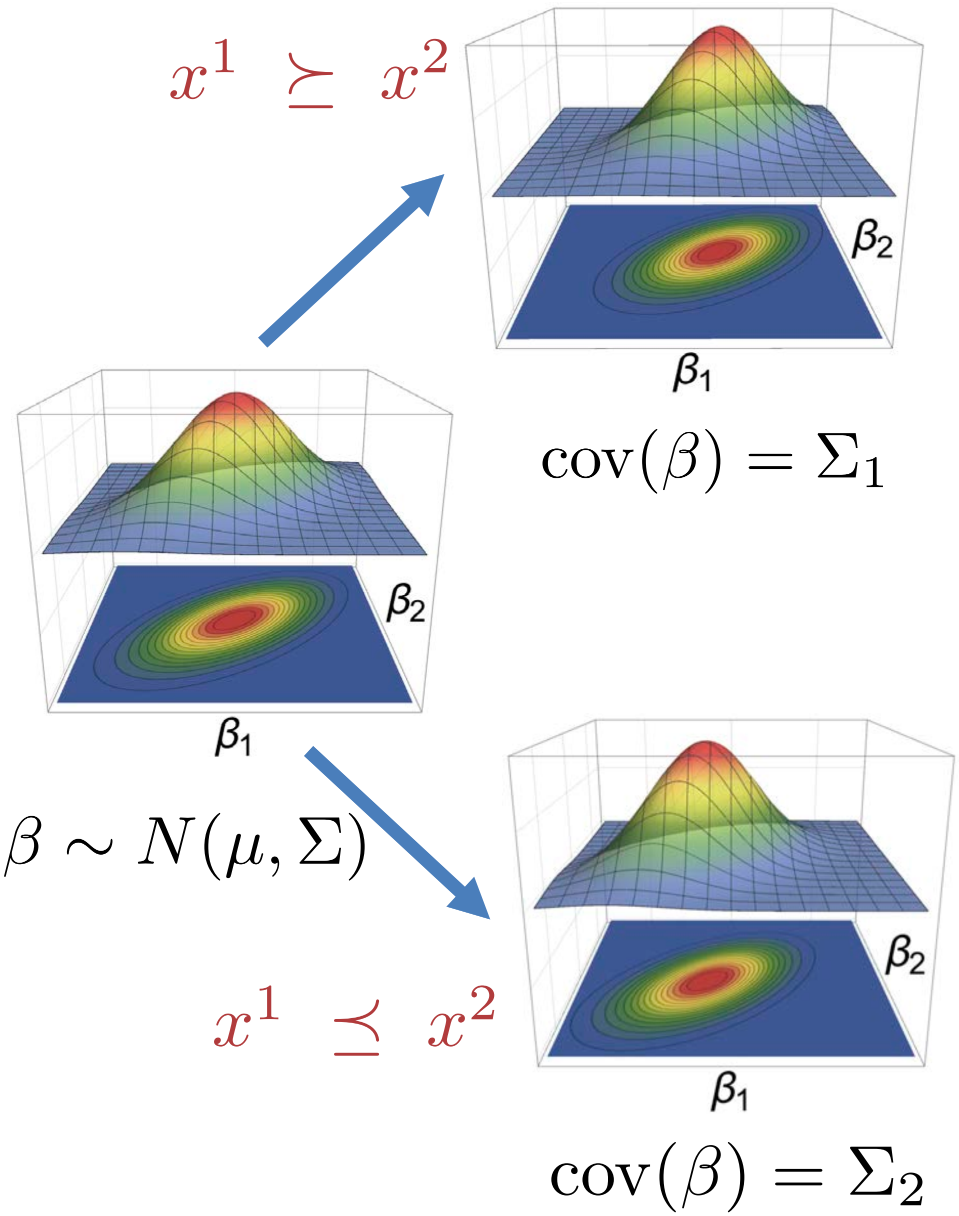
$\operatorname{cov}(\beta$

Only requirements:

- $\beta \sim N(\mu, \Sigma)$
- $L(Y \mid \beta, Z) = \prod_{i=1}^{q} h\left(y^i, \beta \cdot z^i\right)$

✓ Logistic regression with small q

Question Selection for CBCA



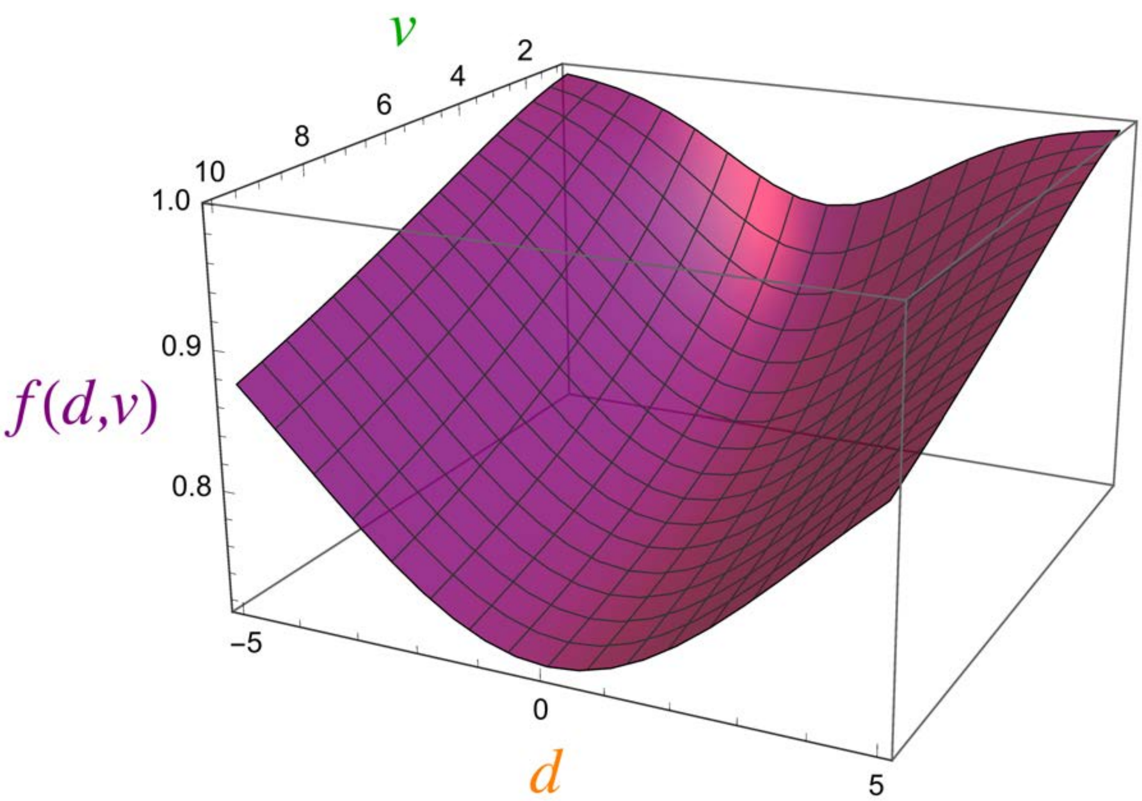
- "Variance" = D-Efficiency:
- $f\left(x^1, x^2\right) := \mathbb{E}_{\beta, x^1 \preceq/\succeq x^2}\left(\det(\Sigma_i)^{1/p}\right)$
- Non-convex function
- Without previous slide, even evaluation requires $\dim(\beta)$ - dimensional integration
-

D-efficiency Simplification for CBCA

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

  distance: $d := \mu \cdot (x^1 - x^2)$

  variance: $v := (x^1 - x^2)' \cdot \sum \cdot (x^1 - x^2)$



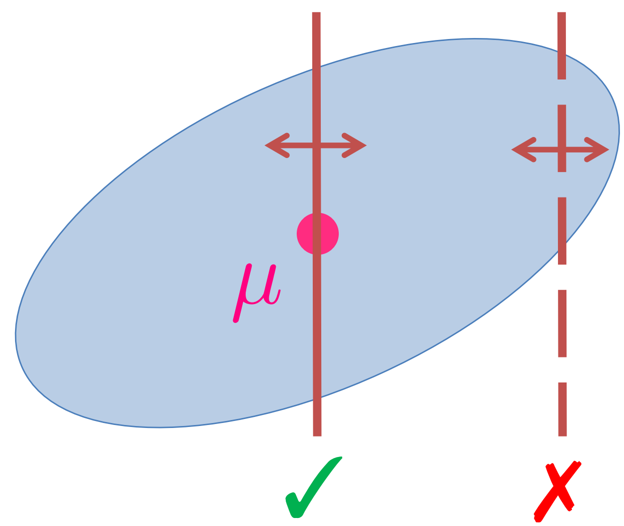
Can evaluate $f(d, v)$ with 1-dim integral 🙂

Simplification = Trade-off for known criteria

$$(\beta - \mu)' \cdot \Sigma^{-1} \cdot (\beta - \mu) \leq r$$

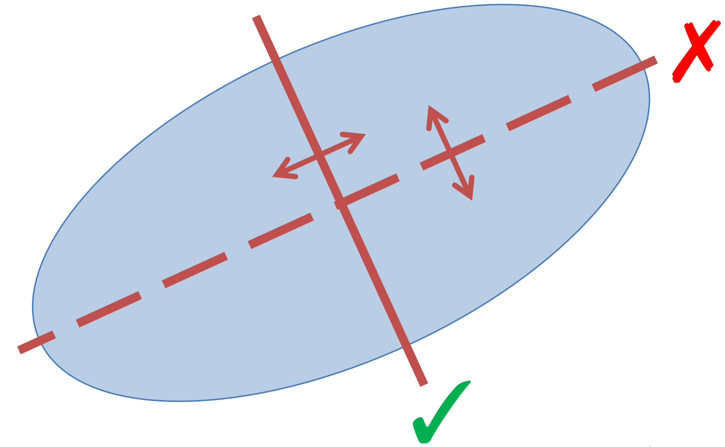
- Choice balance:
  - Minimize distance to center

$$\mu \cdot \left(x^1 - x^2\right)$$



- Postchoice symmetry:
  - Maximize variance of question

$$\left(x^1 - x^2\right)' \cdot \sum \cdot \left(x^1 - x^2\right)$$



Optimization Model

$$\min \quad f(d, v)$$ ✗

$$s.t.$$

$$\mu \cdot (x^1 - x^2) = d$$ ✓

$$(x^1 - x^2)' \cdot \Sigma \cdot (x^1 - x^2) = v$$ ✗ ✓

$$A^1 x^1 + A^2 x^2 \leq b$$ ✓

$$x^1 \neq x^2$$ ✗ ✓

linearize $x_i^k \cdot x_j^l$

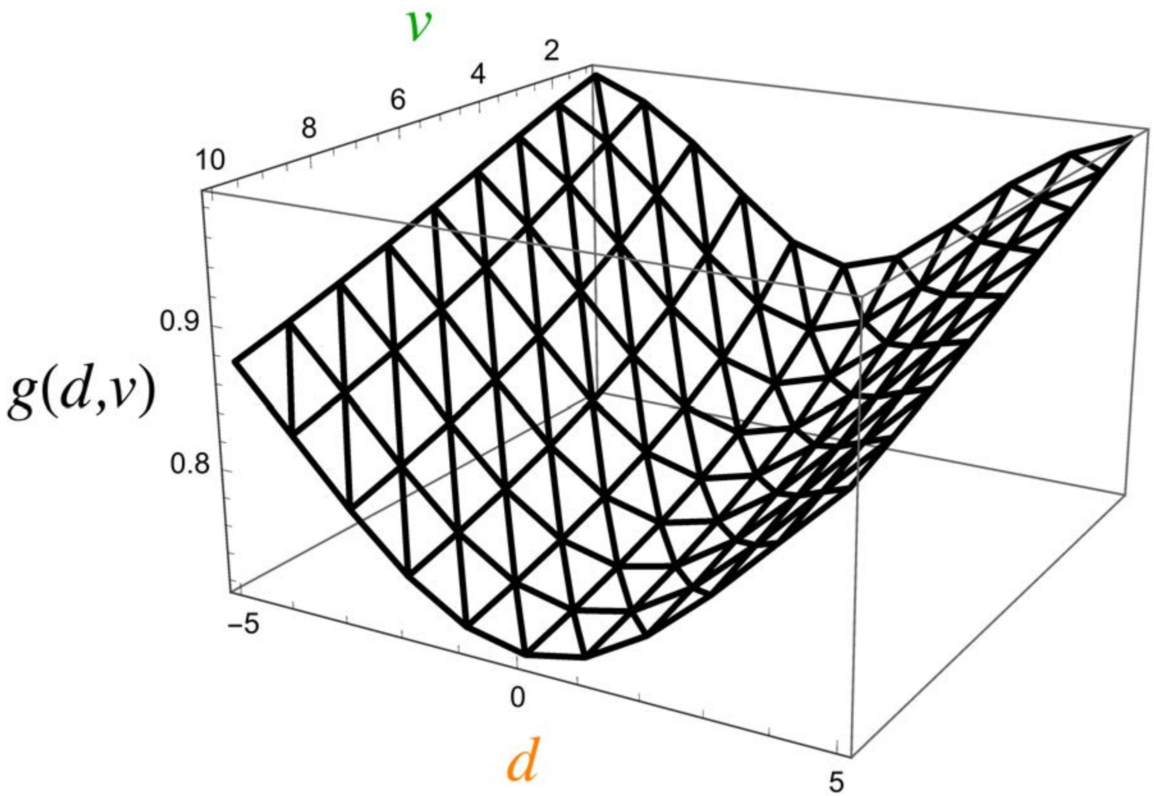
$$x^1, x^2 \in \{0, 1\}^n$$

Technique 2: Piecewise Linear Functions

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

$$\text{distance:} \quad d := \mu \cdot (x^1 - x^2)$$

$$\text{variance:} \quad v := (x^1 - x^2)' \cdot \sum \cdot (x^1 - x^2)$$



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

Piecewise Linear Interpolation

MIP formulation

MIP-based Adaptive Questionnaires

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	☑	☐

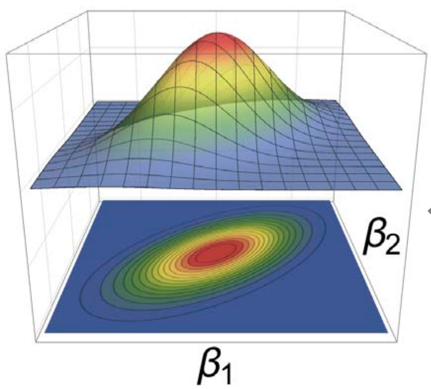
$$\mathbb{E}\left(\beta \mid Y, X^1, X^2\right)$$

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

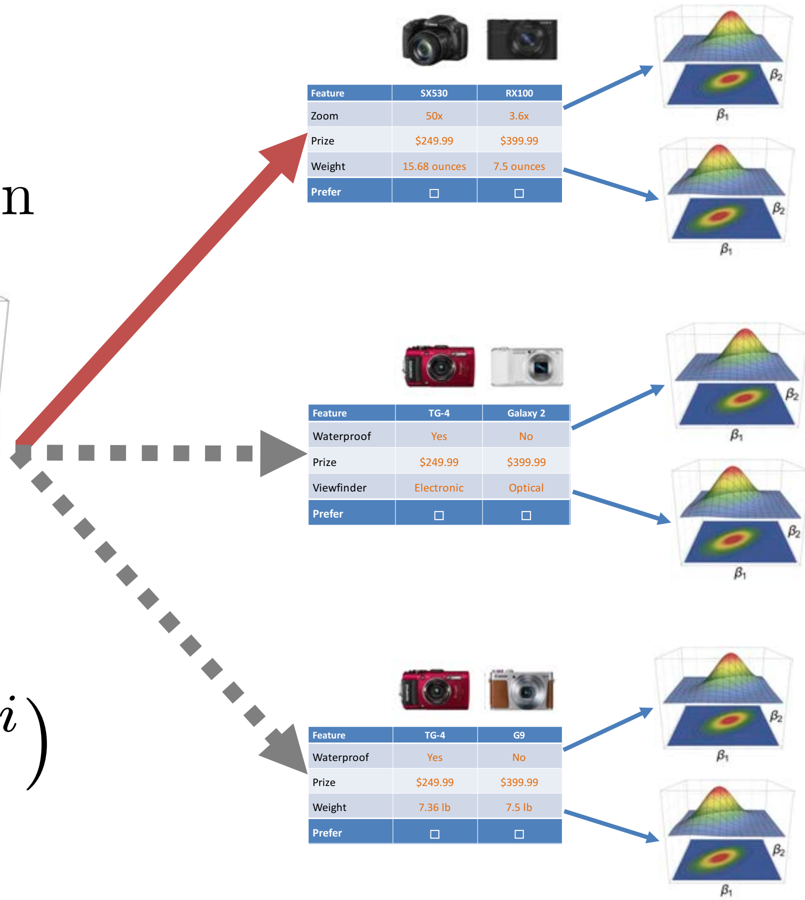
- Optimal one-step look-ahead moment-matching approximate Bayesian approach.

Optimal One-Step Look-Ahead

Prior distribution



$$\beta \sim N\left(\mu^i, \Sigma^i\right)$$



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

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

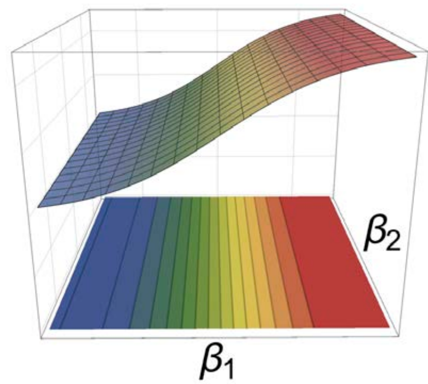
Feature	TG-4	G9
Waterproof	Yes	No
Prize	\$249.99	\$399.99
Weight	7.36 lb	7.5 lb
Prefer	☐	☐

$$\min_{x^1, x^2} f\left(x^1, x^2\right)$$

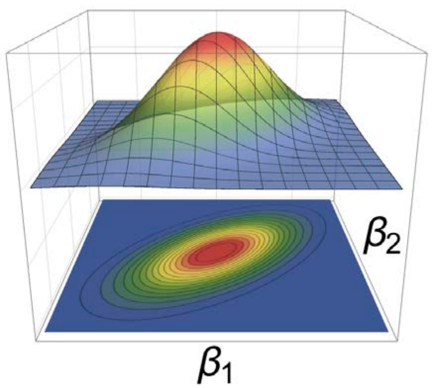
- Solve with MIP formulation

Moment-Matching Approximate Bayesian Update

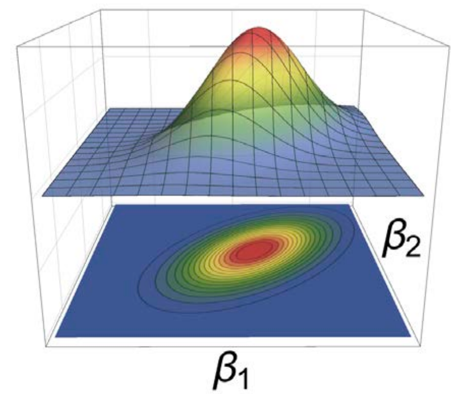
Answer likelihood



Prior distribution



Posterior distribution



Feature	TG-4	Galaxy 2
Waterproof	Yes	No
Prize	\$249.99	\$399.99
Viewfinder	Electron	Optical
Prefer	✓	☐

$$\beta \sim N\left(\mu^i, \Sigma^i\right)$$

$$\beta \overset{approx.}{\sim} N\left(\mu^{i+1}, \Sigma^{i+1}\right)$$

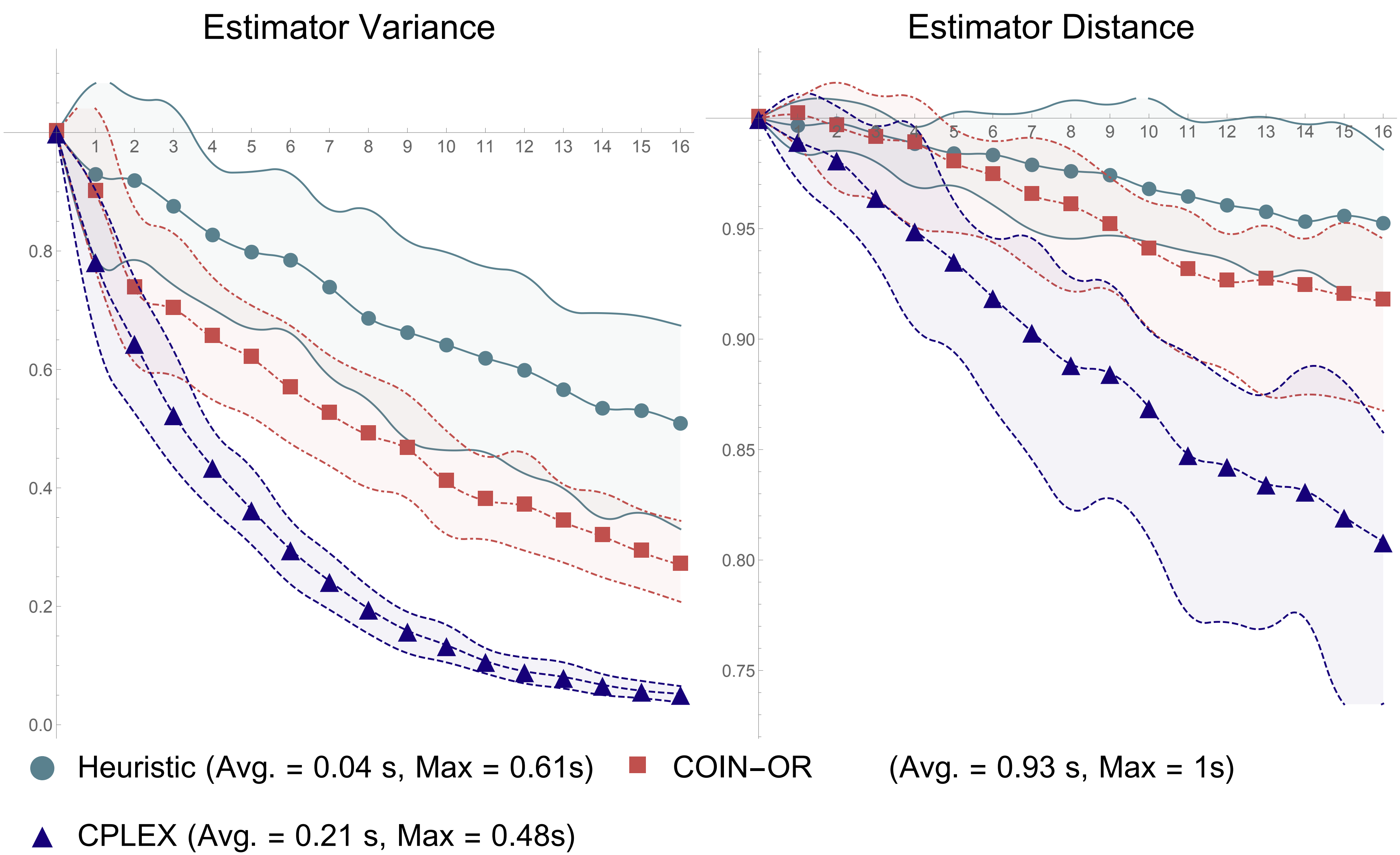
- $\mu^{i+1} = \mathbb{E}\left(\beta \,|\, y, x^1, x^2\right)$
- $\Sigma^{i+1} = \text{cov}\left(\beta \,|\, y, x^1, x^2\right)$

1-dim integral

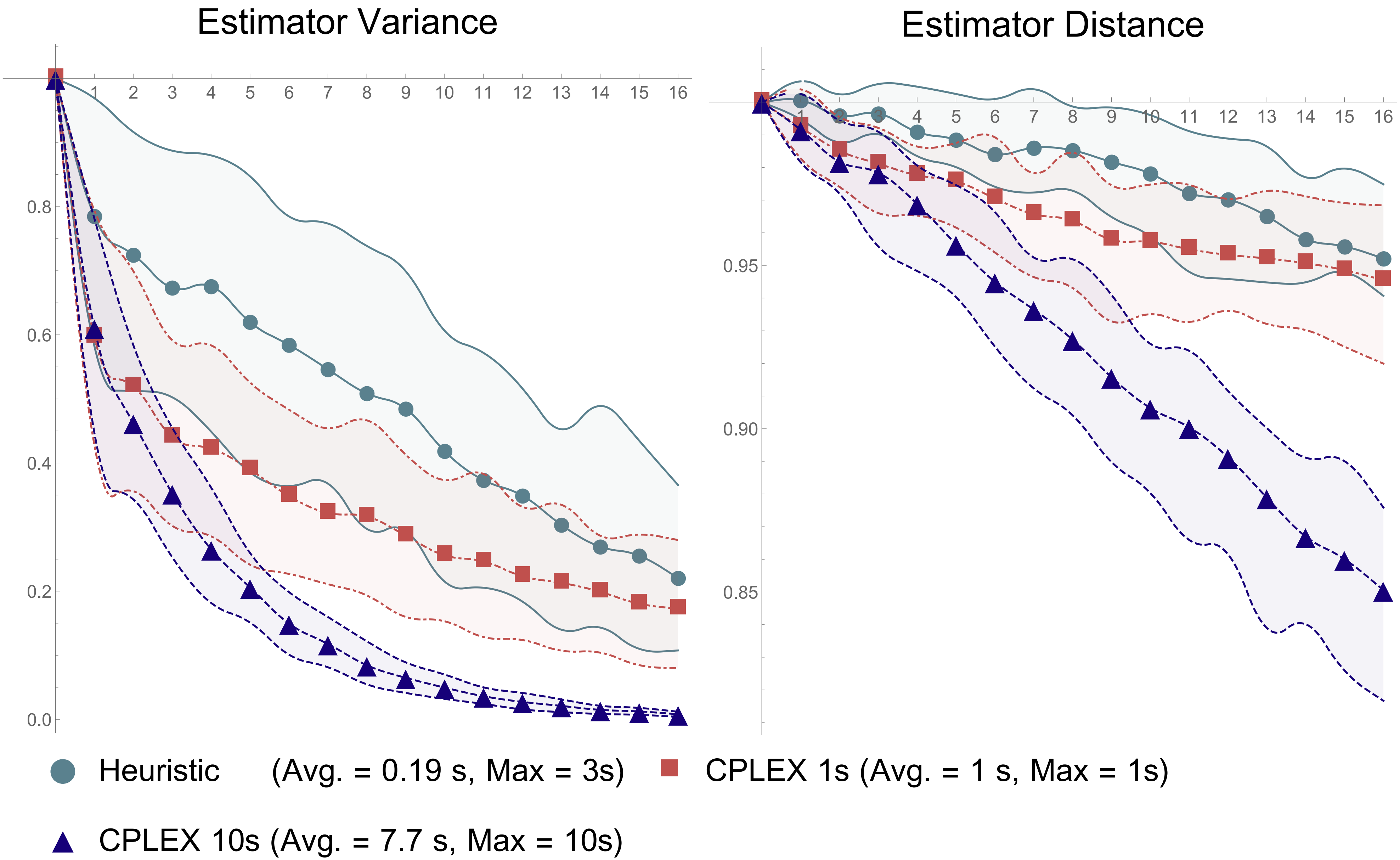
Computational Experiments

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

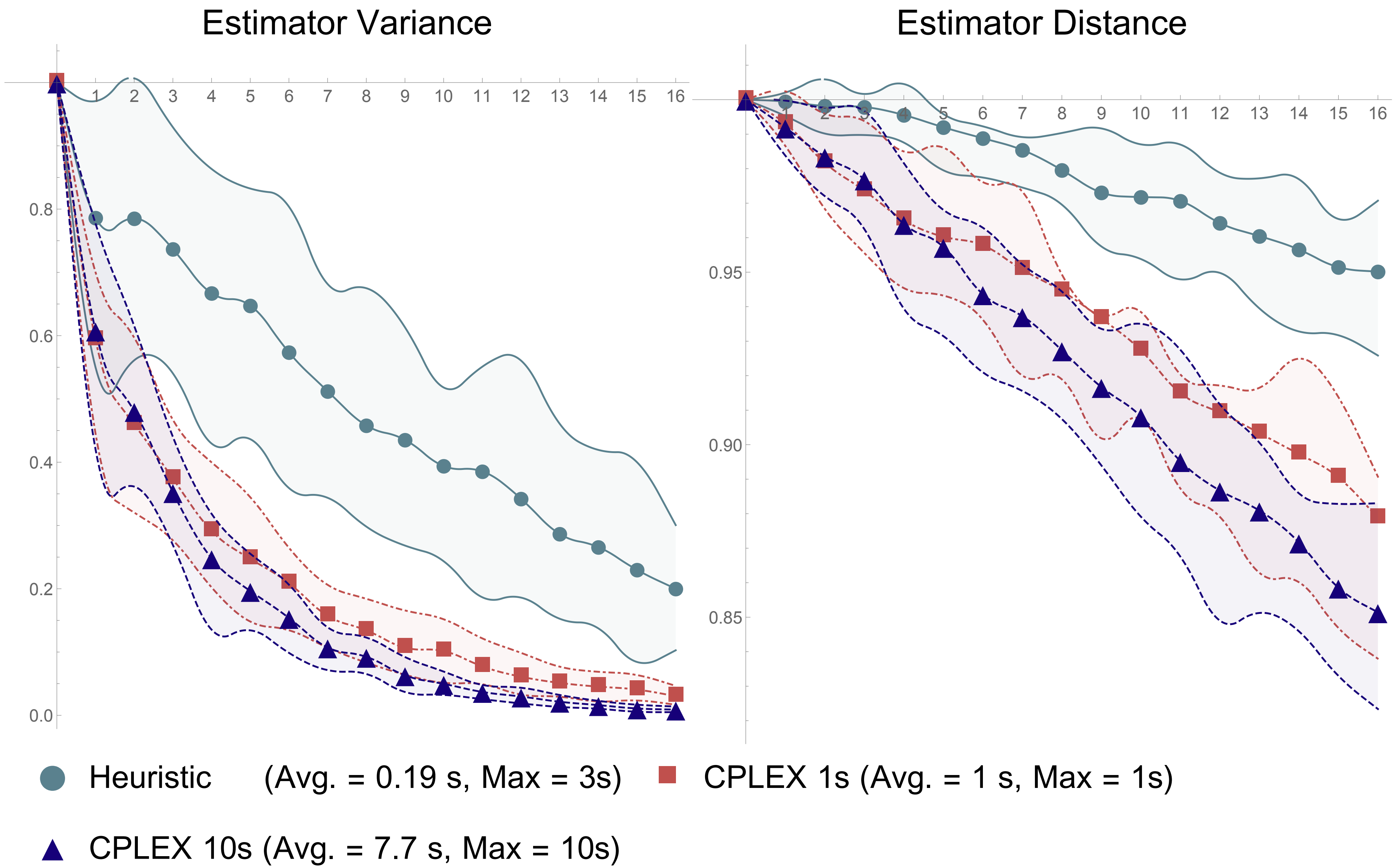
Results for 12 Features, 1 s time limit



Does it Scale? Results for 24 features



Some improvements for 24 features



Summary and Main Messages

- Always choose Chewbacca!
- MIP can now “solve” challenging problems in practice
  - Even in 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
- Adaptive Choice-based Conjoint Analysis
  - Improves on existing **geometric** methods
  - http://ssrn.com/abstract=2798984