Using Julia and JuMP for personalized product recommendations: Ellipsoidal methods for adaptive choice-based conjoint analysis

Juan Pablo Vielma

Massachusetts Institute of Technology

Joint work with Denis Saure

Universidad Adolfo Ibañez,

Santiago, Chile. August, 2017.

JuMP & julia

Hope I am Preaching to the Choir!

Julia and JuMP Tutorial at Universidad Adolfo Ibáñez, Santiago, Chile. January, 2014.

```
sandwich = [:italiano,:chacarero]
@defVar(m, 0 <= P[sandwich] <= 1)
```

21st Century Programming/Modelling Languages

julia

- Open-source and free!
- Developed at MIT
- "Floats like python/matlab, stings like C/Fortran"
- Easy to use and wide library ecosystem (specialized and frontend)
- Only language besides C/C++/Fortran to scale to 1 Petaflop!

JuMP

- Open-source and free!
- Developed at

OPERATIONS RESEARCH CENTER

MIT MANAGEMENT SLOAN SCHOOL

- Optimization modelling language and interphase
- Easy to use and advanced
- Integrated into Julia

JuMP

Created by students

Iain Dunning, Miles Lubin and Joey Huchette

Community Developers

Google Research

OPERATIONS RESEARCH CENTER

Software Engineer

Jarrett Revels

MIT MANAGEMENT SLOAN SCHOOL

Advisor/Boss/$$ … JuMP-Suit?

Godfather?

Juan Pablo Vielma

MIT MANAGEMENT SLOAN SCHOOL

Product Recommendation, Choice-Based Conjoint Analysis, and Experimental Design

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:

Towards CBCA-Based Recommendations

- Individual preference estimates with few questions
- Adaptive Questions:
  - Fast question selection
  - Pick **next** question to reduce uncertainty
  - Quantify estimate variance
- Favorable properties for future:
  - Intuitive geometric model (e.g. Robust Opt.)
  - Parametric model

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

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

Parametric Model = Logistic Regression

- MNL Random **Linear** Utilities (d product features)

$$U_p = \beta \cdot x^p + \epsilon_p = \sum_{i=1}^{d} \beta_i x_i^p + \epsilon_p \quad p \in \{1, 2\}$$

product profile — part-worths (weights) — noise (gumbel)

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

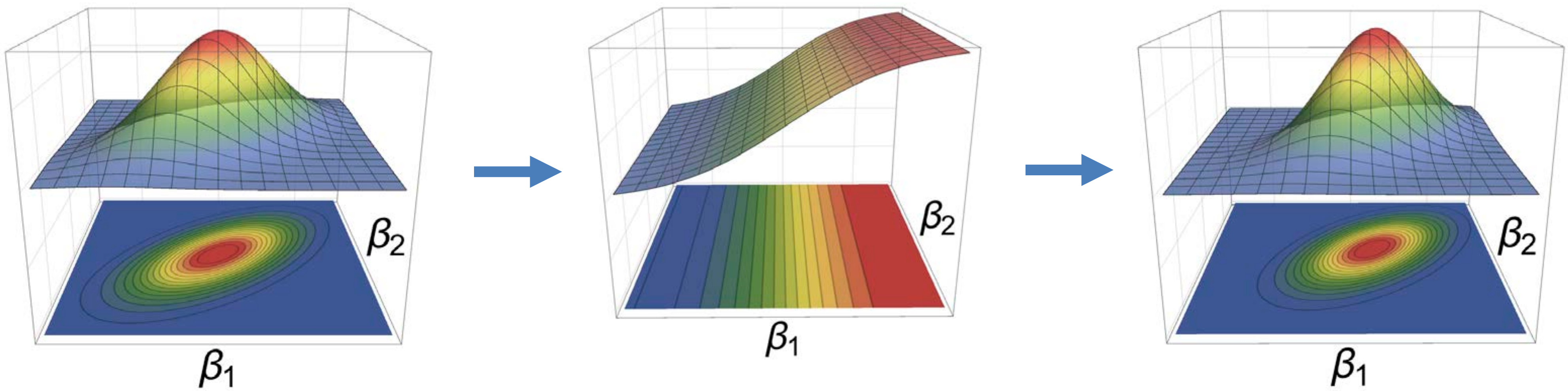
- Noise = response error: $\mathbb{P}\left(x^1 \succeq x^2 \mid \beta\right) = \dfrac{1}{1 + e^{-\beta \cdot (x^1 - x^2)}}$

- Regression: $z := x^1 - x^2, \quad y := \text{sign}(\beta \cdot z) \in \{0, 1\}$

$$x^1 \succeq x^2 \Leftrightarrow \beta \cdot z \text{"}\geq\text{"} 0 \Leftrightarrow \text{sign}(\beta \cdot z) = 1$$

Bayesian Update After a Question is Answered

Prior distribution | Answer likelihood | Posterior distribution



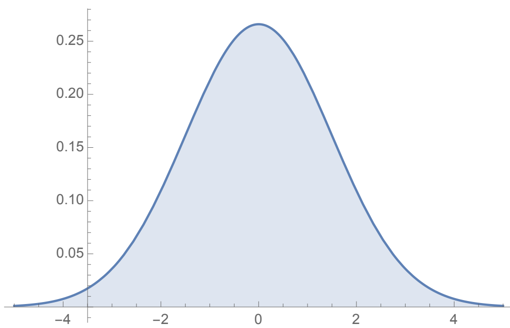
$$\beta \sim N(\mu, \Sigma) \qquad L\,(y \mid \beta, z) \qquad g\,(\beta \mid y, z\,)$$

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

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

Pick Next Question To Reduce Posterior "Variance"

Prior Distribution of β



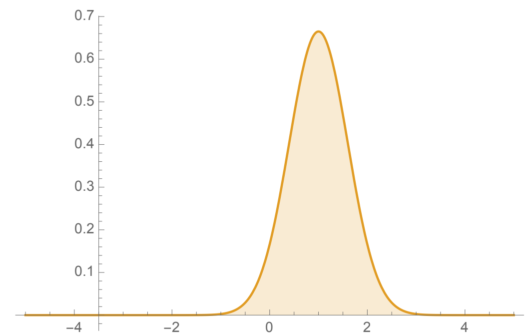
$\beta \sim N(\mu, \Sigma)$

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	☑	☐

Bayesian Update

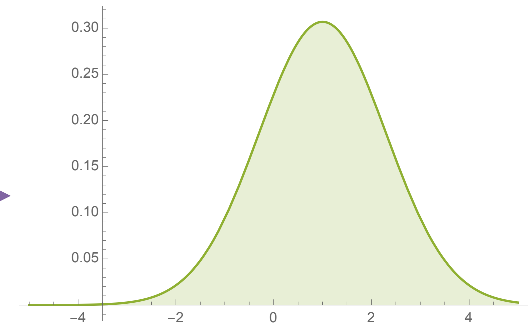
Posterior Distribution



MCMC

Bayesian Update

Posterior Distribution



- Multivariate version with uncertainty in answer: D-Error:

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

Bayesian D-Optimal Question Selection

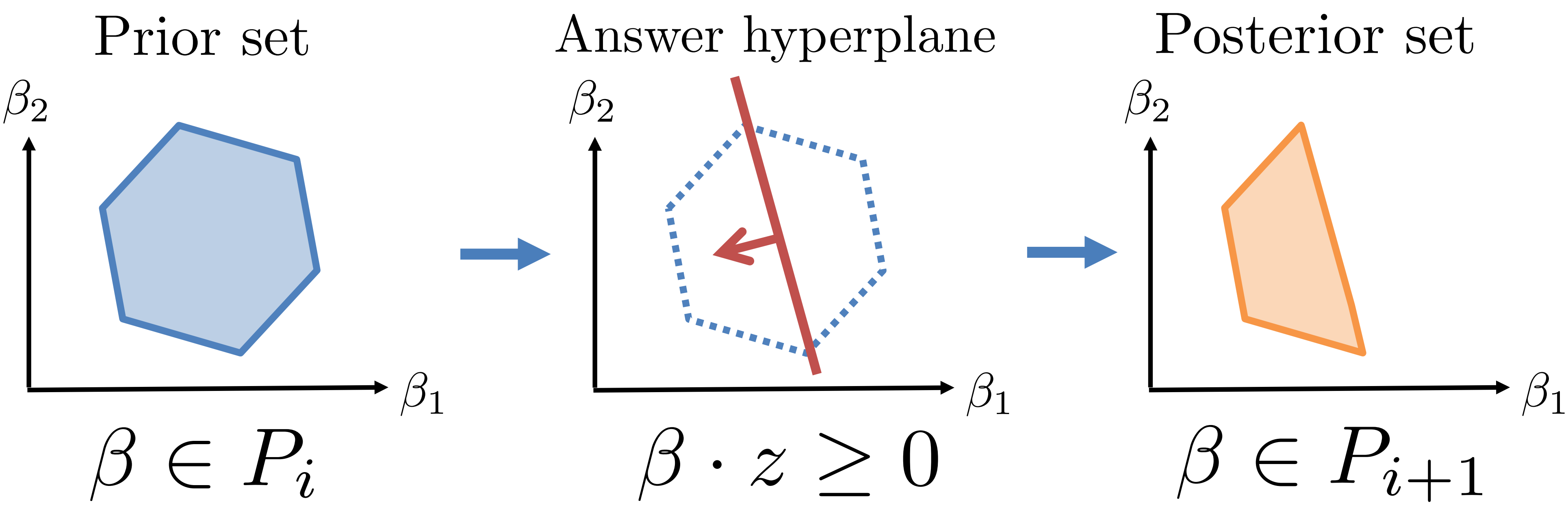
State-of-the-art MCMC tools easy to access from Julia (Stan.jl):

Minutes to evaluate one question for 12 features = **Months** to find best!

"Fisher Information approximations":

- **Minutes** to find **"best"** by enumeration
- High dimensional non-convex optimization.
- Low/High variance numerical issues.

Alternative: Fast Geometric/Optimization Models



Method	Response Error
Polyhedral Method (Toubia et al. ‘03,’04)	No
Probabilistic Polyhedral Method (T. et al. ‘07)	Yes, ≈ Bayesian
Robust Method (Bertsimas and O’Hair ‘13)	Yes, Robust

Bayesian v/s Geometric

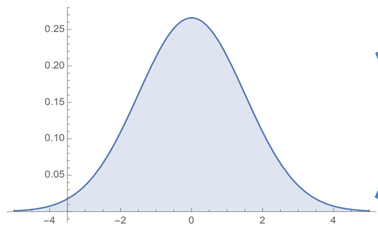
	Bayesian	Geometric
Response Error	MNL	None / Non-MNL
Update	Integration or MCMC	Simple Linear Algebra
Question Selection	Integration + Enumeration	MIP

- **Ellipsoidal Method**:
  - MIP, julia and JuMP bridges the GAP (**Question Selection**)

Mixed Integer Programming and Optimal Question Selection

Next Question: Reduce “Variance” / D-Error

Prior Distribution of β



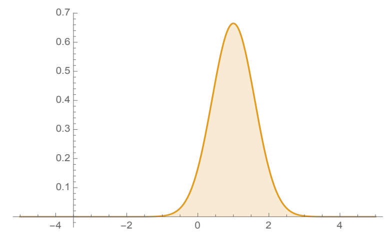
$$\beta \sim N(\mu, \Sigma)$$

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

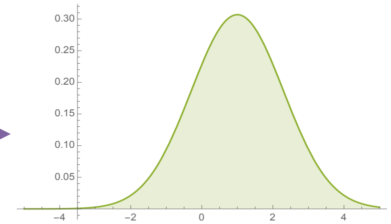
Feature	TG-4	Galaxy 2
Waterproof	Yes	No
Prize	$249.99	$399.99
Viewfinder	Electronic	Optical
Prefer	✓	☐

Update

Posterior Distribution



Posterior Distribution



$$\min_{x^1 \neq x^2 \in \{0,1\}^d} f\left(x^1 - x^2, \mu, \Sigma\right)$$

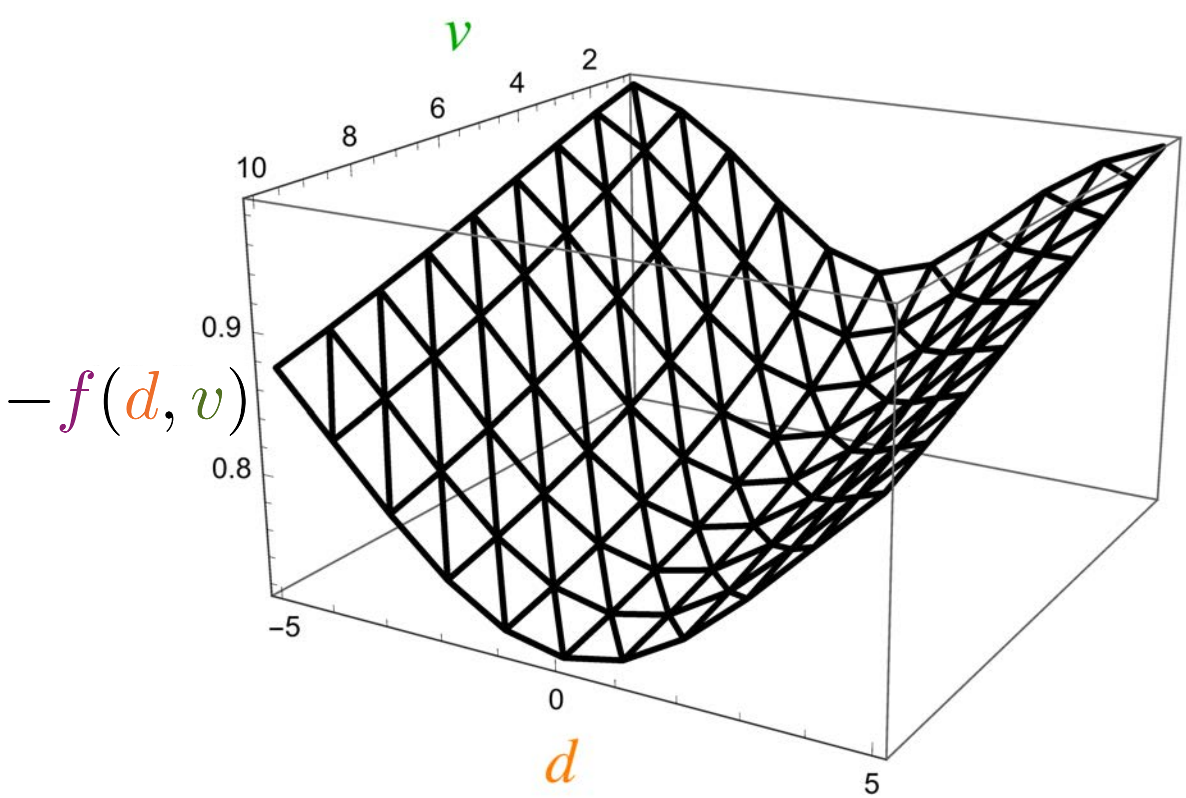
$$f(z, \mu, \Sigma) := \mathbb{E}_{y,\beta}\left\{(\det \operatorname{cov}(\beta \mid y, z))^{1/d}\right\}$$

Low Dimensional Reformulation of D-Error

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

mean: $d := \mu \cdot z$

variance: $v := z' \cdot \Sigma \cdot z$



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

Piecewise Linear (PWL) Interpolation

Linear MIP formulation (standard linearization)

Aligns with selection criteria from Toubia et al. '04: minimize mean and maximize variance

Linear MIP: Linearize Quad + PWL Formulation

$$\min \quad f(d, v)$$

$$s.t.$$

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

$$\left(x^1 - x^2\right)' \cdot \Sigma \cdot \left(x^1 - x^2\right) = v$$

$$A^1 x^1 + A^2 x^2 \leq b$$

$$x^1 \neq x^2$$

linearize $x_i^k \cdot x_j^l$

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

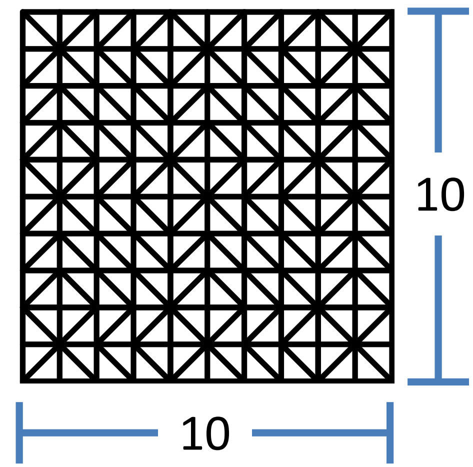
Easy to Build through julia & JuMP

- PiecewiseLinearOpt.jl (Huchette and V. 2017)

$$
\begin{aligned}
\min \quad & \exp(x+y) \\
s.t. \quad & \\
& x, y \in [0,1]
\end{aligned}
$$

Automatically select Δ

Automatically construct formulation (easily chosen)




```
using JuMP, PiecewiseLinearOpt
m = Model()
@variable(m, x)
@variable(m, y)

z = piecewiselinear(m, x, y, 0:0.1:1, 0:0.1:1, (u,v) -> exp(u+v))
@objective(m, Min, z)
```

Ellipsoidal Method:
Putting Everything Together

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	☑	☐

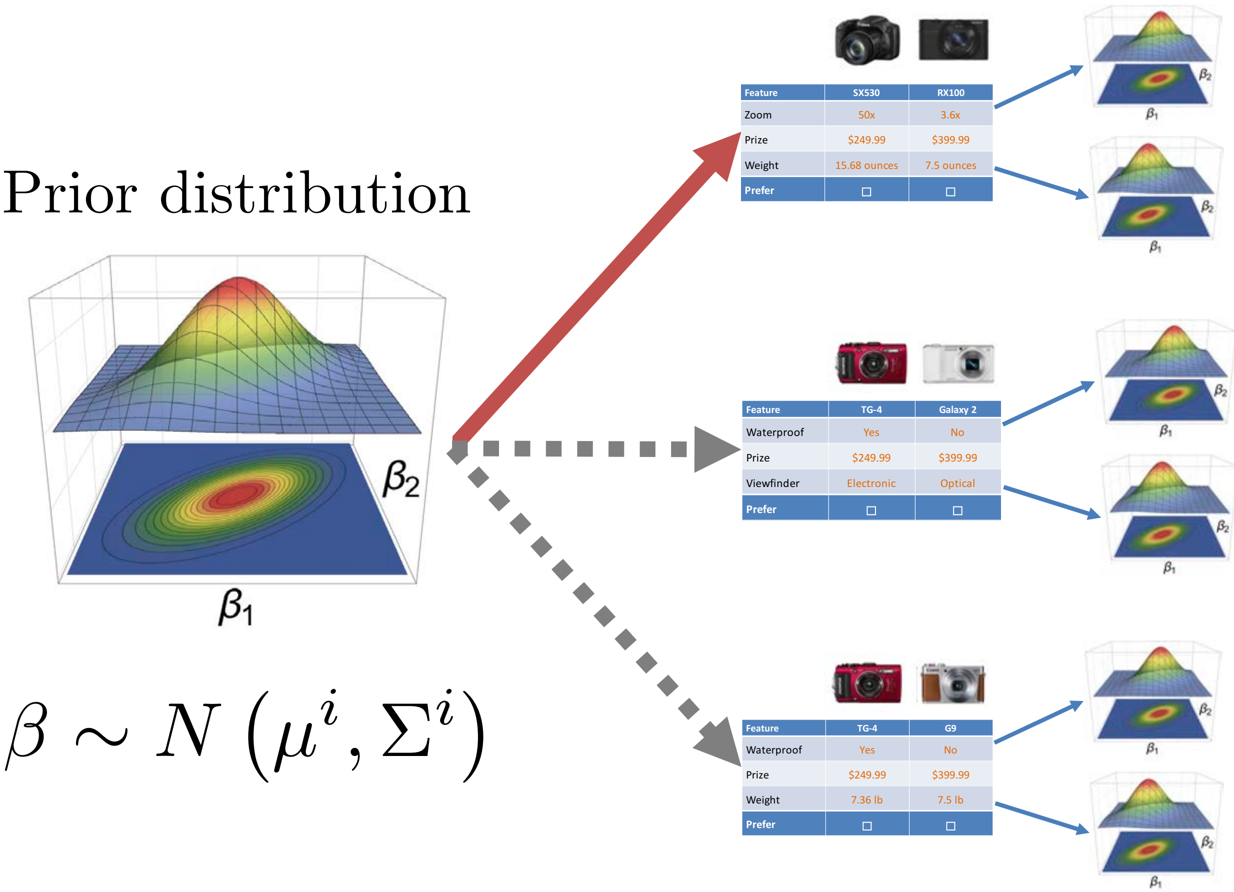
$$\mathbb{E}\left(\beta \,|\, Y, X^1, X^2\right)$$

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

- **Optimal one-step look-ahead moment-matching approximate Bayesian approach = Ellipsoidal Method**

Optimal One-Step Look-Ahead = MIP

Prior distribution



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

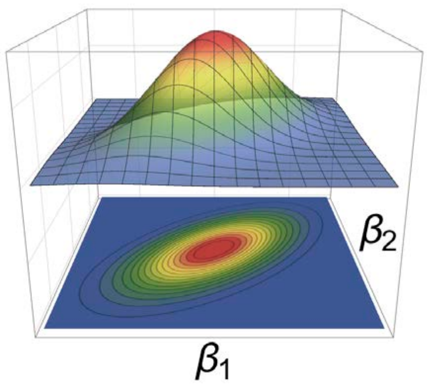
$$\beta \sim N\left(\mu^i, \Sigma^i\right)$$

$$\min_{x^1, x^2, d, v \in Q} f(d, v)$$

- Solve with MIP formulation
- 1-dim numerical integration: QuadGK.jl

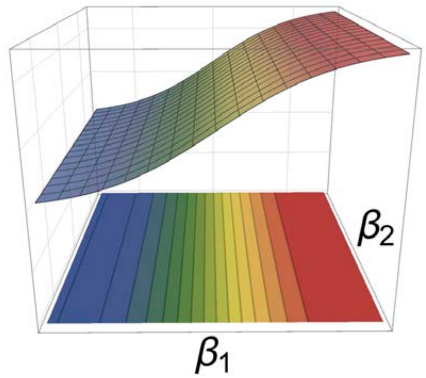
Moment-Matching Approximate Bayesian Update

Prior distribution



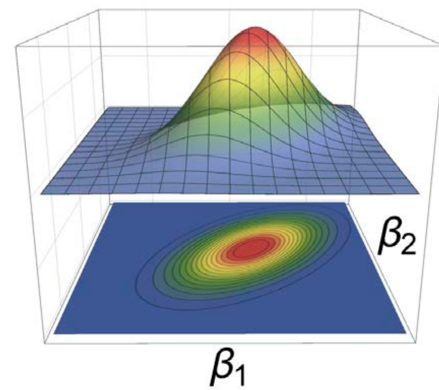
$$\beta \sim N\left(\mu^i, \Sigma^i\right)$$

Answer likelihood



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

Posterior distribution



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

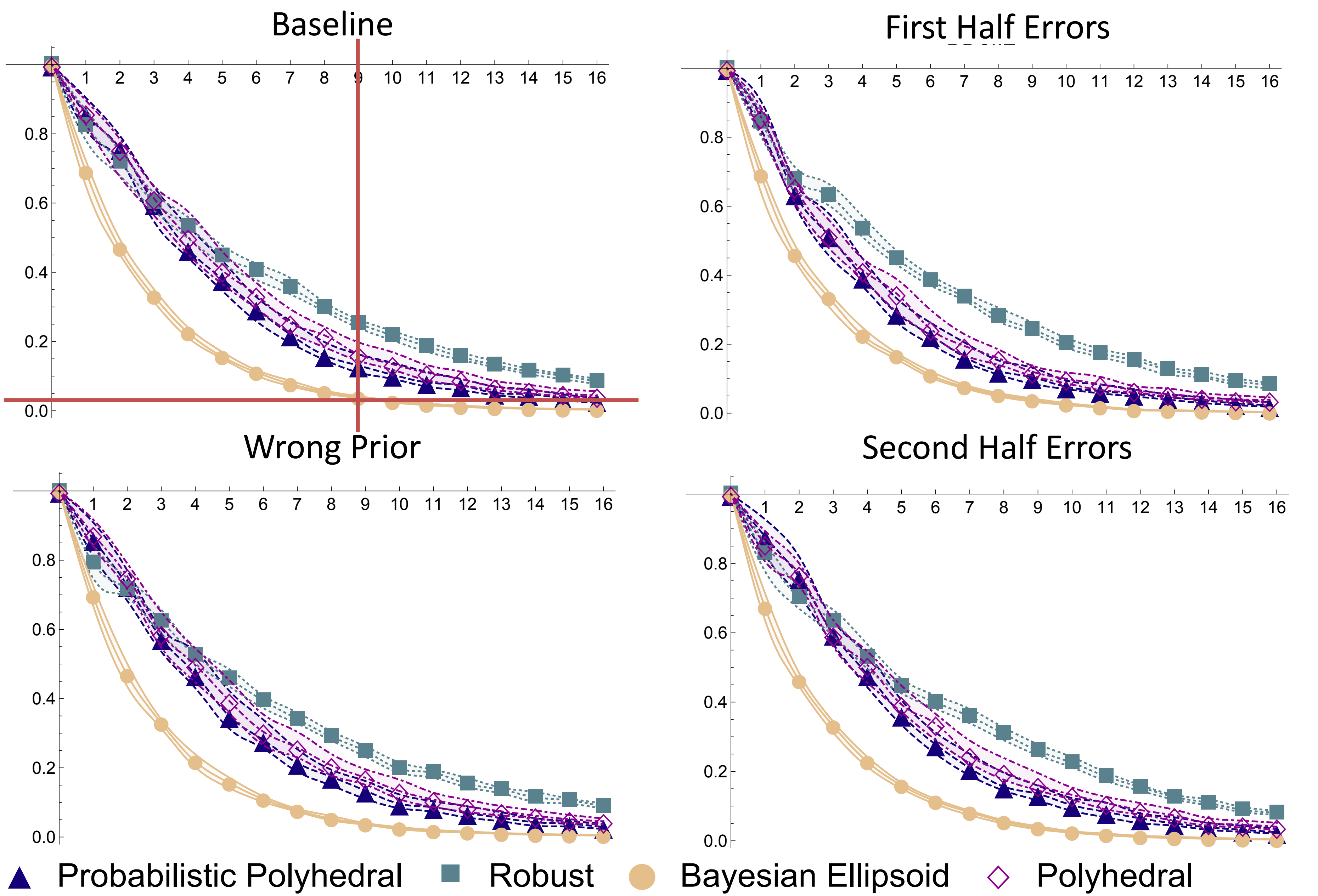
- $\mu^{i+1} = \mathbb{E}\left(\beta \,|\, y, x^1, x^2\right)$
- $\Sigma^{i+1} = \mathrm{cov}\left(\beta \,|\, y, x^1, x^2\right)$

- 1-d integral : $I(d, v)$

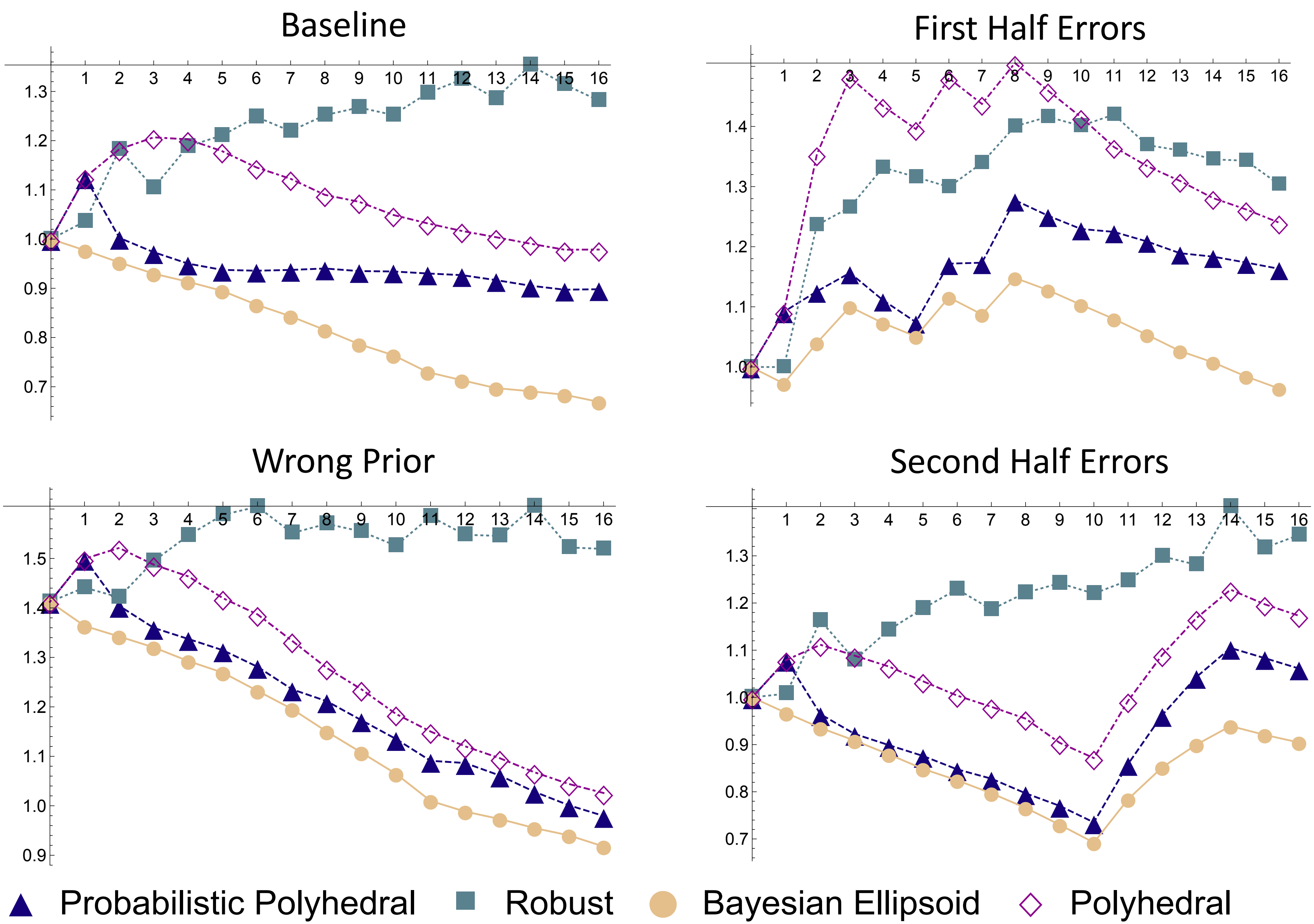
Simulation Experiments

- 16 questions, 2 options, 12 features
- Simulate MNL responses with known β^*
- 100 individual β^* sampled from N(μ,Σ) prior
- Methods:
  - Polyhedral, Prob. Polyhedral, Robust and Ellipsoidal
  - All get same ellipsoidal prior
  - All < 30’’ inter-question (except robust < 90’’)
- Metrics:
  - RMSE of β estimator, error in market share and D-eff.
  - Normalized values = smaller better
  - Versions: Method and Bayesian Estimator
  - Sensitivity: Wrong prior μ, all errors in first/second half

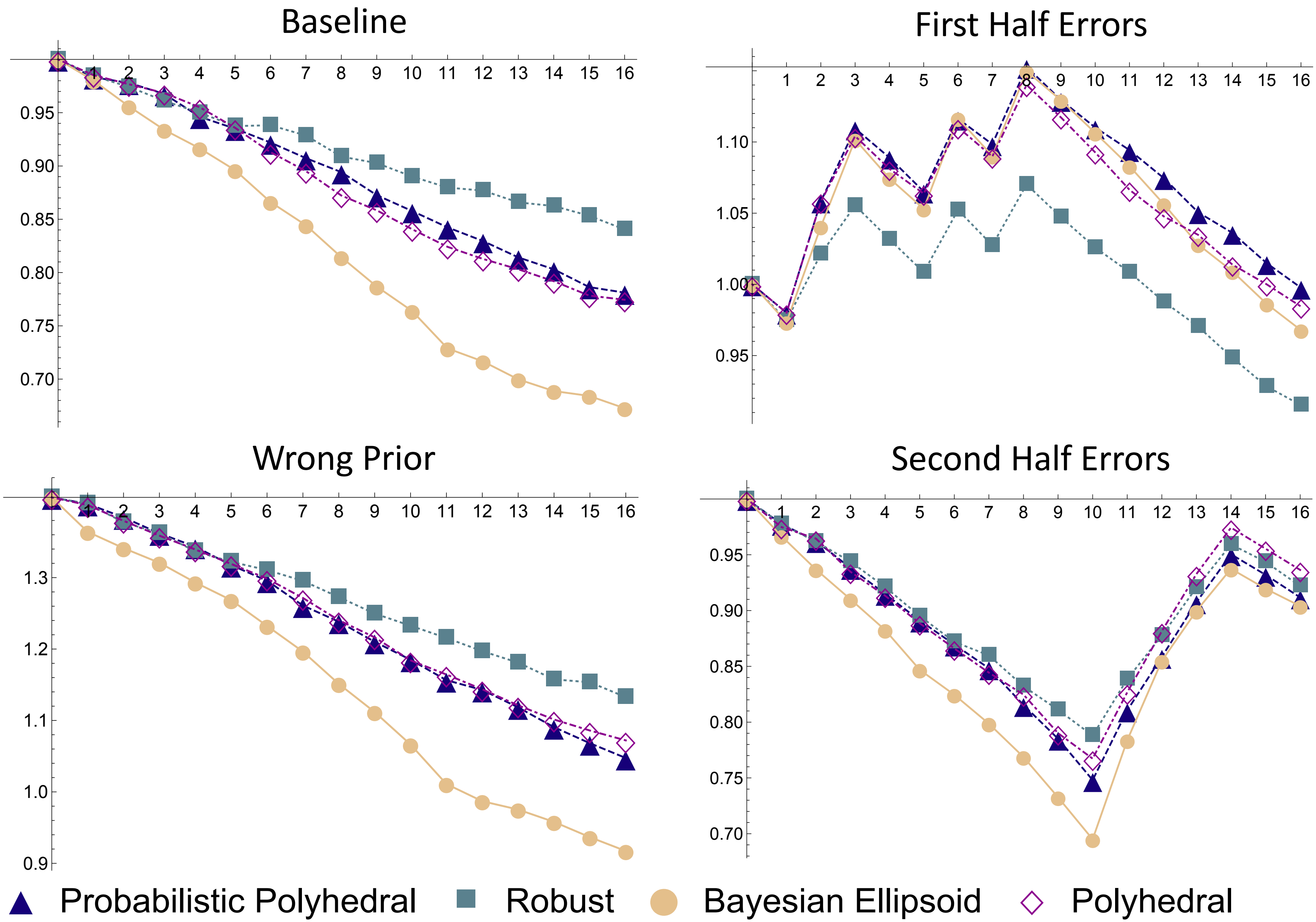
D-Efficiency for Individual Bayesian



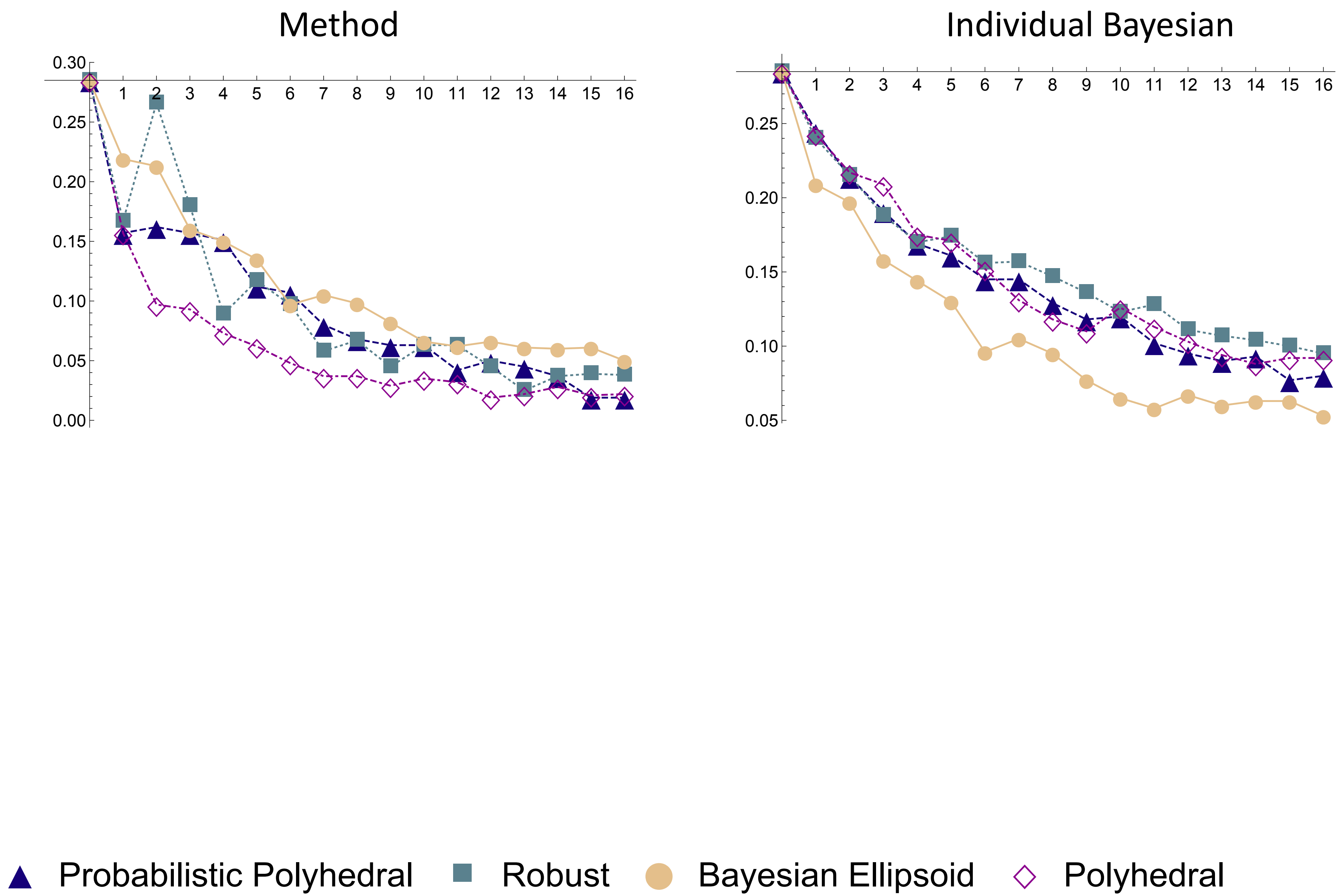
RMSE for Methods Estimator



RMSE for Individual Bayesian Estimator



Market share for Baseline



Summary

- Mixed Integer Programming for ACBCA
  - n-variate function to 2-variate function + MIP
  - Advanced MIP formulation + solver
  - Easy to access with JuMP !
- Also for other estimator variance / linear models
- Significantly faster reduction of estimator variance
- Future:
  - Julia Package
  - MIP flexibility → Managerial Objective