Mixed Integer Programming Approaches for Real-Time Consumer Preference Elicitation (and Causal Inference)

Juan Pablo Vielma
Massachusetts Institute of Technology

Operations Management/Management Science Workshop
Booth School of Business, University of Chicago.
Chicago, IL, October, 2017.
(Nonlinear) Mixed Integer Programming (MIP)

\[ \min \quad f(x) \]
\[ \text{s.t.} \quad x \in C \]
\[ x_i \in \mathbb{Z} \quad i \in I \]
Mostly convex \( f \) and \( C \).

Marketing And Experimental Design

Causal Inference for Educational Impact of 2010 Chilean Earthquake

“Infinite”-Dimensional MIP and Control or Aerial Drones

http://www.gurobi.com/company/example-customers
50+ Years of MIP = Significant Solver Speedups

- Algorithmic Improvements (*Machine Independent*):
  - Commercial Solver Speedup \( \approx 1.9 \times \text{year} \)

- Mostly linear, but also quadratic:
  - Gurobi v6.0 (2014) – v6.5 (2015) quadratic: \( 4.43 \times \)
    *(V., Dunning, Huchette, Lubin, 2015)*

- Also great “open-source” solvers

  - SCIP
  - CBC
  - GLPK

- Emerging: General Convex Nonlinear (e.g. SDP)

  - KNITRO
  - Bonmin
MIP Modelling and Advanced Formulations

• MIP Representability: What can be modeled with MIP?
  – Linear: Jeroslow & Lowe ‘80s ... Basu, Martin, Ryan and Wang ’17
  – Convex Nonlinear:
    • MIP formulation for the set of Prime Numbers
      ✓ Non-Convex Polynomial MIP formulation (Jones et al. ’76)
      ✗ Convex of any kind (Lubin, Zadik and V. ’17)

• Linear/nonlinear formulation techniques:
Accessing Solvers = Modelling Languages

- User-friendly algebraic modelling languages (AML):
  - Standalone and Fast
  - Based on General Language and Versatile

- 21st Century AML:
  - Free and Open-Source
  - Easy to use, but as advanced as proprietary C/C++ interphases
  - As fast as standalone AMLs and C/C++ interphases
Created by students

Community Developers

Software Engineer

JuMP-Suit?

Iain Dunning, Miles Lubin
and Joey Huchette
2016 ICS Prize

Juan Pablo Vielma
Outline

• MIP and Consumer Preference Elicitation
  – Direct improvement from MIP formulation
  – Fast, versatile and efficient learning

• MIP and Causal Inference
  – Indirect improvement from MIP formulation
  – Right formulation brings you back to solving the problem you really wanted to solve
Mixed Integer Programming
(joint work with Joey Huchette)

and

Consumer Preference Elicitation
(joint work with Denis Saure)
Adaptive Choice-Based Conjoint Analysis

<table>
<thead>
<tr>
<th>Feature</th>
<th>SX530</th>
<th>RX100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom</td>
<td>50x</td>
<td>3.6x</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Weight</td>
<td>15.68 ounces</td>
<td>7.5 ounces</td>
</tr>
<tr>
<td>Prefer</td>
<td>✓</td>
<td>❑</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>TG-4</th>
<th>Galaxy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterproof</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Weight</td>
<td>7.36 lb</td>
<td>7.5 lb</td>
</tr>
<tr>
<td>Prefer</td>
<td>❑</td>
<td>✓</td>
</tr>
</tbody>
</table>

Today: Minimize variance of parameter estimates

Estimate of preference parameter
Parametric Model = Logistic Regression

<table>
<thead>
<tr>
<th>Feature</th>
<th>Chewbacca</th>
<th>BB-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wookiee</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Droid</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Blaster</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prefer?</td>
<td>☑</td>
<td>☐</td>
</tr>
</tbody>
</table>

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

Product profile

\[ U_j = \beta \cdot x^j + \epsilon_j \]

MNL Random Linear Utility

\[ \sum_{j=1}^{d} \beta_i x_i^j \]

Question:

\[ x^1 \succ x^2 \iff U_1 \text{ "} > \text{"} U_2 \]

\[ \iff \beta \cdot z \text{ "} > \text{"} 0 \]

\[ \mathbb{P} \left( x^1 \succ x^2 \mid \beta \right) = \frac{1}{1 + e^{-\beta \cdot z}} \]
Bayesian Model with Normal Prior

Prior distribution

\[ \beta \sim N(\mu, \Sigma) \]

Answer likelihood

\[ L(y \mid \beta, z) \]

Posterior distribution

\[ g(\beta \mid y, z) \]

\[ y = \text{sign}(\beta \cdot z) \]

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

\[ g(\beta \mid y, z) \propto \phi(\beta \mid \mu, \Sigma) L(y \mid \beta, z) \]
D-Error and Expected Posterior Variance

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

\[ \text{cov}(\beta) = \Sigma_1 \]

\[ \text{cov}(\beta) = \Sigma_2 \]

\[ y < 0 \quad y > 0 \]

\[ \beta \sim \mathcal{N}(\mu, \Sigma) \]

\[ \min_{z \in \{-1,0,1\}^n \setminus \{0\}} f(z, \mu, \Sigma) \]

\[ f(z, \mu, \Sigma) \text{ is hard to evaluate, non-convex and } n \text{ large} \]
1st Step: Moment-Matching Approximate Bayes

Prior distribution

\[ \beta \sim N (\mu^i, \Sigma^i) \]

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

Answer likelihood

Posterior distribution

\[ \beta^{\text{approx.}} \sim N (\mu^{i+1}, \Sigma^{i+1}) \]

- Linear Algebra + 1-d numerical integration (e.g. BDA3)

---

<table>
<thead>
<tr>
<th>Feature</th>
<th>TG-4</th>
<th>Galaxy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterproof</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Viewfinder</td>
<td>Electronic</td>
<td>Optical</td>
</tr>
<tr>
<td>Prefer</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>
2nd Step: More Linear Algebra from V. and S. ‘16

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

mean: \( d := \mu \cdot z \)

variance: \( v := z' \cdot \sum \cdot z \)

Can evaluate \( f(d, v) \) with 1-dim integral 😊

Piecewise Linear (PWL) Interpolation \( \hat{f}(d, v) \)

Balances known criteria:
- minimize mean of question (no clear expected answer)
- maximize variance of question (uncertainty in expected answer)
3rd Step: “Almost” Direct Linear MIP Formulation

\[ z = x^1 - x^2 \]

MIP formulation for PWL function

\[ \min \hat{f}(d, v) \]

s.t.

\[ \mu \cdot (x^1 - x^2) = d \]

\[ (x^1 - x^2)' \cdot \sum \cdot (x^1 - x^2) = v \]

linearize \( x_i^k \cdot x_j^l \) \( \|x^1 - x^2\|_2^2 \geq 1 \) \( (x^1 \neq x^2) \)

\[ x^1, x^2 \in \{0, 1\}^n \]
Simple Formulation for Univariate Functions

\[ z = f(x) \]

\[
\begin{pmatrix}
  x \\
  z
\end{pmatrix} = \sum_{j=1}^{5} \begin{pmatrix}
  d_j \\
  f(d_j)
\end{pmatrix} \lambda_j
\]

\[ 1 = \sum_{j=1}^{5} \lambda_j, \quad \lambda_j \geq 0 \]

\[ y \in \{0, 1\}^4, \quad \sum_{i=1}^{4} y_i = 1 \]

\[ 0 \leq \lambda_1 \leq y_1 \]
\[ 0 \leq \lambda_2 \leq y_1 + y_2 \]
\[ 0 \leq \lambda_3 \leq y_2 + y_3 \]
\[ 0 \leq \lambda_4 \leq y_3 + y_4 \]
\[ 0 \leq \lambda_5 \leq y_4 \]

Size = \( O(\# \text{ of segments}) \)

Non-Ideal: Fractional Extreme Points
Advanced Formulation for Univariate Functions

\[ z = f(x) \]

\[ \begin{pmatrix} x \\ z \end{pmatrix} = \sum_{j=1}^{5} \begin{pmatrix} d_j \\ f(d_j) \end{pmatrix} \lambda_j \]

\[ 1 = \sum_{j=1}^{5} \lambda_j, \quad \lambda_j \geq 0 \]

\[ y \in \{0, 1\}^2 \]

\[ 0 \leq \lambda_1 + \lambda_5 \leq 1 - y_1 \]

\[ 0 \leq \lambda_3 \leq y_1 \]

\[ 0 \leq \lambda_4 + \lambda_5 \leq 1 - y_2 \]

\[ 0 \leq \lambda_1 + \lambda_2 \leq y_2 \]

Size = \( O(\log_2 \# \text{ of segments}) \)

Ideal: Integral Extreme Points

Significant computational advantage

V. and Nemhauser, 2011.
Technique Also Works for Multivariate Functions

- Union Jack triangulation (V. and Nemhauser, 2011)
  \[ \text{Size} = 4 \lceil \log_2 n \rceil + 2 \]

- For general triangulations (Huchette and V., 2016, 2017)
  \[ \text{Size} \leq 4 \lceil \log_2 n \rceil + 6 \]
  - Based on finding a bi-clique cover of an auxiliary graph
  - Can use a MIP to find the smallest formulation!
• PiecewiseLinearOpt.jl (Huchette and V. 2017)

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

```julia
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)
```

Automatically select \( \Delta \)

Automatically construct formulation (easily chosen)
function getquestion(μ,Σ,variancefuction)
    n = size(Σ,1)
    m = Model()
    # define variables for linearization
    @variable(m, 0 <= x[1:n] <= 1, Int)
    @variable(m, 0 <= y[1:n] <= 1, Int)
    # x ≠ y
    @constraint(m, linquad(m,(x-y)·(x-y)) >= 1)
    # v = x-y, β ~ N(μ,Σ), v·β ~ N(μv,σ²), μv = μ·v, σ² = v′Σ*ν
    @variable(m, μv)
    @constraint(m, μv == μ·(x-y) )
    @variable(m, σ² >=0)
    @constraint(m, σ² == linquad(m,(x-y)·(Σ*(x-y))))
    # (x-y)′Σ*(x-y) <= eigmax(Σ) ||x-y||₂ <= eigmax(Σ)*n
    σ² = eigmax(Σ)*n
    # (x-y)′Σ*(x-y) >= eigmin(Σ) ||x-y||₂ >= eigmin(Σ) ( x ≠ y )
    σ² = eigmin(Σ)
    μv = norm(μ,1)
    μvnpoints = 2^k - 1
    μvpoints = 0:μv/μvnpoints:μv+(μv/μvnpoints)/2
    σ²range = σ² - σ²
    σ²npoints = 2^k-1
    σ²points = σ²:σ²range/σ²npoints:σ²+(σ²range/σ²npoints)/2
    pwl = BivariatePWLFunction(μvpoints, σ²points, (μv,σ²) -> variancefuction(μv,sqrt(σ²)))
    obj = piecewiselinear(m, μv, σ², pwl)
    @objective(m, Min, obj )
    status = solve(m)
    return [ round(Int64,getValue(x)), round(Int64,getValue(y))]
MIP v/s Best Benchmark (Toubia et al. ‘03,’04)

- 16 questions, 2 options, 12 features, 100 individual $\beta^*$ sampled from known prior $N(\mu, \Sigma)$
- Best benchmark v/s MIP + Moment Matching
- CPLEX: $\leq 1$ s (0.2 s Avg.), GLPK: $\leq 5$ s (1.7 s Avg.)
Easy To Add Questionnaire Rules

<table>
<thead>
<tr>
<th>Feature</th>
<th>Chewbacca</th>
<th>BB-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wookiee</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Droid</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Blaster</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prefer?</td>
<td>✓</td>
<td>□</td>
</tr>
</tbody>
</table>

Product profile

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

- Realism is important: Wookiees are not Droids!
  \[
x^1_{\text{Wookie}} + x^1_{\text{Droid}} \leq 1
\]

- Partial Profiles:
  - Limit # of feature differences and assume those not shown are the same (e.g. both are members of the resistance)
  \[
  \| x^1 - x^2 \|_1 \leq 3
  \]
Full v/s Partial Profiles ( # Feature Differences )

- 16 questions, 2 options, 12 features, 100 individual $\beta^*$ sampled from known prior $\mathcal{N}(\mu, \Sigma)$
- Best benchmark v/s MIP + Moment Matching (CPLEX)
- Full: $\leq 1$ s (0.2 s Avg.), Partial (5 diff.): $\leq 66$ s (8 s Avg.)
MIP and Causal Inference

Joint work with Magdalena Bennett and Jose Zubizarreta
Educational Impact of 2010 Chilean Earthquake

• 6th Strongest in Recorded History (8.8)

• Impact on Educational Achievement (PSU Scores)?
Very High Quality Data is Available

Case study: impact of an earthquake on educational outcomes

Intensity of the earthquake

José R. Zubizarreta (Columbia)

Matching using Integer Programming

Test Scores and Demographic Info
Covariate Balance Important for Inference

- **Dose 1** = Control = Affected by Earthquake
- **Dose 2** = Treatment = Not affected by Earthquake

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Dose 1</th>
<th>Dose 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>462</td>
<td>462</td>
</tr>
<tr>
<td>Female</td>
<td>538</td>
<td>538</td>
</tr>
<tr>
<td>School SES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Mid-low</td>
<td>327</td>
<td>327</td>
</tr>
<tr>
<td>Medium</td>
<td>294</td>
<td>294</td>
</tr>
<tr>
<td>Mid-high</td>
<td>189</td>
<td>189</td>
</tr>
<tr>
<td>High</td>
<td>115</td>
<td>115</td>
</tr>
<tr>
<td>Mother’s education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>335</td>
<td>335</td>
</tr>
<tr>
<td>Secondary</td>
<td>426</td>
<td>426</td>
</tr>
<tr>
<td>Technical</td>
<td>114</td>
<td>114</td>
</tr>
<tr>
<td>College</td>
<td>114</td>
<td>114</td>
</tr>
<tr>
<td>Missing</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>
Traditional Matching: Exact v/s Fine Balance

**Exact Matching**
Match units with same category in all covariates

**Fine Balance**
Different matches for different covariates
Matching v/s MIP

- **Matching**: 
  \[ \# \text{ Variables} = (\# \text{ treated}) \times (\# \text{ controls}) \]

- **Simple MIP**: 
  - Just count & balance units in each category/covariate
  - \[ \# \text{ Variables} = (\# \text{ treated}) + (\# \text{ controls}) \]
  - Can use known results (Balas and Pulleyblank, 1983) to show MIP formulation is as strong as matching formulation
  - Can show MIP formulation is integral for 2 covariates
  - Problem is NP-hard for > 2 covariates
  - Usually very fast solve times: cuppa coffee time \( \approx 5 \) min
  - More flexible.... doses and representability!!
Multiple Doses + Representation

- Base, Medium and High

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Dose</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td>462</td>
<td>462</td>
<td>462</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>538</td>
<td>538</td>
<td>538</td>
</tr>
<tr>
<td>School SES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td>75</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Mid-low</td>
<td></td>
<td>327</td>
<td>327</td>
<td>327</td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td>294</td>
<td>294</td>
<td>294</td>
</tr>
<tr>
<td>Mid-high</td>
<td></td>
<td>189</td>
<td>189</td>
<td>189</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>115</td>
<td>115</td>
<td>115</td>
</tr>
<tr>
<td>Mother’s education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td></td>
<td>335</td>
<td>335</td>
<td>335</td>
</tr>
<tr>
<td>Secondary</td>
<td></td>
<td>426</td>
<td>426</td>
<td>426</td>
</tr>
<tr>
<td>Technical</td>
<td></td>
<td>114</td>
<td>114</td>
<td>114</td>
</tr>
<tr>
<td>College</td>
<td></td>
<td>114</td>
<td>114</td>
<td>114</td>
</tr>
<tr>
<td>Missing</td>
<td></td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

- Whole population

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Population</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father’s education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>0.39</td>
<td>0.40</td>
</tr>
<tr>
<td>Technical</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>College</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>Missing</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Mother’s education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>0.41</td>
<td>0.43</td>
</tr>
<tr>
<td>Technical</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>College</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>Missing</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Household income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100-200</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>200-400</td>
<td>0.30</td>
<td>0.31</td>
</tr>
<tr>
<td>400-600</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>600-1400</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td>1400 or more</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Missing</td>
<td>0.01</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Template + Multiple Doses

Target population

Representative template sample

Dose 1

Dose 2

Dose D
Relative (To no Quake) Attendance Impact [%]

95% confidence interval

Base
Medium
High
• The interesting question is Why?
  – After class programs? Government support?
Summary

• Advances in MIP
  – Advanced Formulations
  – Advanced Solvers
  – Easy Access Through **JuMP**

• Direct advantage (Choice Based Conjoint Analysis)
  – Real-time and versatile adaptive questionnaires
  – Cut number of questions in half
  – 20% improvement in estimated parameter quality
  – Market-share predictions cut in half

• Indirect advantage (Causal Inference)
  – Good and flexible formulations can bring you back to solving the problem you really wanted to solve