Heuristic and metaheuristic algorithms for the generation of optimal experimental designs

Daniel Palhazi Cuervo
Supervisors: Peter Goos and Kenneth Sörensen

June 15th, 2015 – Antwerp, Belgium
Context and motivation
Correlation does not imply causation!

Math doctorates awarded
- Suicides by hanging, strangulation and suffocation

Degrees awarded
- Suicides in the US

Year
- 1999
- 2000
- 2001
- 2002
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009

600 700 800 900 1,000 1,100 1,200 1,300 1,400 1,500 1,600 1,700 1,800

5,000 5,500 6,000 6,500 7,000 7,500 8,000 8,500 9,000 9,500

600 800 1,000 1,200 1,400 1,600 1,800 2,000 2,200 2,400 2,600 2,800 3,000 3,200 3,400 3,600 3,800 4,000 4,200 4,400 4,600 4,800 5,000 5,200 5,400 5,600 5,800 6,000 6,200 6,400 6,600 6,800 7,000 7,200 7,400 7,600 7,800 8,000 8,200 8,400 8,600 8,800 9,000 9,200 9,400 9,600 9,800 10,000

1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
Experimentation

Controlled input (factors)

Uncontrolled input (covariates)

Output (responses)

\[ y_i = f_i(x_1, x_2, \ldots, x_F) + \epsilon_i \]
Design of experiments (simple example)
Design of experiments (simple example)

True relationship
Design of experiments (simple example)
Design of experiments (simple example)
Design of experiments (simple example)
Design of experiments (simple example)

- True relationship
- Experiment
- Regression model
Design of experiments (simple example)
Design of experiments (simple example)
Design of experiments (simple example)
Design of experiments (simple example)
Design of experiments (baking bread)
### Design of experiments (baking bread)

<table>
<thead>
<tr>
<th>Firmness</th>
<th>Flour</th>
<th>Sugar</th>
<th>Salt</th>
<th>Oil</th>
<th>Oven</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100 g</td>
<td>500 g</td>
<td>2 tbs</td>
<td>1 tbs</td>
<td>200°, 15 min</td>
</tr>
<tr>
<td>2</td>
<td>80 g</td>
<td>700 g</td>
<td>1 tbs</td>
<td>3 tbs</td>
<td>180°, 20 min</td>
</tr>
<tr>
<td>3</td>
<td>130 g</td>
<td>600 g</td>
<td>3 tbs</td>
<td>2 tbs</td>
<td>220°, 25 min</td>
</tr>
<tr>
<td>4</td>
<td>30 g</td>
<td>500 g</td>
<td>1 tbs</td>
<td>1 tbs</td>
<td>240°, 20 min</td>
</tr>
<tr>
<td>8</td>
<td>80 g</td>
<td>600 g</td>
<td>3 tbs</td>
<td>2 tbs</td>
<td>180°, 15 min</td>
</tr>
</tbody>
</table>

\[
y = f(x_1, x_2, x_3, x_4, x_5, x_6) + \varepsilon
\]
Approaches for the design of experiments

**Standard designs**

- Limited set of scenarios
  - Very good properties
  - Experimenters adapt the scenario to fit a design
- Available in catalogues

**Optimal design of experiments**

- Flexible approach
- Best design for the scenario at hand
- Optimization problem (algorithms)
  - Choice of the factor settings
  - Objective function
- Not always achieve the quality of standard designs
Main goal of the PhD

Optimal design of experiments

relatively simple algorithms

Operations research

(meta)heuristic algorithms
Main goal of the PhD

Optimal design of experiments

Operations research

(meta)heuristic algorithms
Industrial Experiments
Context

- First stage of process analysis
- Relatively simple experiment with several factors

Example (baking bread)
Screening experiments

Motivation

• Coordinate-exchange algorithm (restart strategy)

Contribution

• Classification of existing algorithms
• New selection strategy (orthogonality measure)
  ▪ Reduction in the execution time by 30%
• Iterated local search algorithm
  ▪ More efficient designs than the restart strategy (large experiments)
Two-stratum experiments

Context

• Limit in the number of observations that can be made under homogeneous conditions
• Factors that are hard or expensive to change
### Example (baking bread)

<table>
<thead>
<tr>
<th>Firmness</th>
<th>Flour</th>
<th>Sugar</th>
<th>Salt</th>
<th>Oil</th>
<th>Oven</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>500 g</td>
<td>2 tbs</td>
<td>1 tbs</td>
<td>2 tbs</td>
<td>15 min</td>
</tr>
<tr>
<td>2</td>
<td>700 g</td>
<td>1 tbs</td>
<td>3 tbs</td>
<td>1 tbs</td>
<td>20 min</td>
</tr>
<tr>
<td>3</td>
<td>600 g</td>
<td>3 tbs</td>
<td>2 tbs</td>
<td>3 tbs</td>
<td>25 min</td>
</tr>
<tr>
<td>4</td>
<td>700 g</td>
<td>1 tbs</td>
<td>1 tbs</td>
<td>1 tbs</td>
<td>20 min</td>
</tr>
<tr>
<td>7</td>
<td>500 g</td>
<td>2 tbs</td>
<td>1 tbs</td>
<td>2 tbs</td>
<td>25 min</td>
</tr>
<tr>
<td>8</td>
<td>600 g</td>
<td>3 tbs</td>
<td>2 tbs</td>
<td>3 tbs</td>
<td>15 min</td>
</tr>
</tbody>
</table>
Motivation

- *Traditional approach:* groups of equal size
- *Assumption:* the number of groups and the size of each group are limited only by upper bounds
- *Goal:* to identify the best grouping configuration

Contribution

- Algorithmic framework (multi-level optimization)
  - Low level: factor-level configurations
  - High level: grouping configuration
- For some scenarios, designs with groups of different sizes are up to 9% more efficient
Stated choice experiments
SC experiments with partial profiles

Context
- Trade-off that people make when choosing between competing options

Example (bakery)

<table>
<thead>
<tr>
<th>Factors</th>
<th>Alternative #1</th>
<th>Alternative #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opening hours</td>
<td>6 am – 4 pm</td>
<td>9 am – 6 pm</td>
</tr>
<tr>
<td>Types of bread</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Average price</td>
<td>3.30 euros</td>
<td>2.75 euros</td>
</tr>
<tr>
<td>Sales cakes and pastry</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Open on Sundays</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
**SC experiments with partial profiles**

**Context**
- Trade-off that people make when choosing between competing options

**Example (bakery)**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Alternative #1</th>
<th>Alternative #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opening hours</td>
<td>6 am – 4 pm</td>
<td>9 am – 6 pm</td>
</tr>
<tr>
<td>Types of bread</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Average price</td>
<td>3.30 euros</td>
<td>2.75 euros</td>
</tr>
<tr>
<td>Sales cakes and pastry</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Open on Sundays</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Motivation
• Limited scenarios handled by standard designs
• Weaknesses in existing algorithms

Contribution
• Integrated algorithm (simultaneous optimization)
  ▪ Set of factors that are held constant
  ▪ Levels of the varying factors
• Matches the efficiency of standard designs
• Outperforms the existing algorithms (designs that are up to 11% more efficient)
Motivation

- Lack of literature about algorithmic methods
- Link to designs for main-effects models and partial profiles

Contribution

- Comparison between standard designs and algorithmic methods
  - Algorithmic methods usually generate designs that are more efficient
- Extension of the integrated algorithm for the generation of designs with additional constant attributes
In a nutshell...

The benefits of the optimal design of experiments

1. The experimental designs generated by algorithmic methods can match the quality of the standard designs

2. It is possible to leverage the flexibility provided by these algorithms in order to generate even better designs
Heuristic and metaheuristic algorithms for the generation of optimal experimental designs

Daniel Palhazi Cuervo
Supervisors: Peter Goos and Kenneth Sörensen

June 15th, 2015 - Antwerp