Constrained Optimization with Sample Estimates of Parameters

James J. Cochran
Distinguished Professor of Statistics & Operations Research
Mike & Kathy Mouron Research Chair
The University of Alabama
jjcochran@ua.edu

Applied Math in Statistics and Data Science Education Workshop
Institute for Computational and Experimental Research in Mathematics
Brown University
Providence RI
Tuesday, June 10, 2025

“The artist sees what others only catch a glimpse of.”

“L'artista vede ciò che gli altri riescono solo a intravedere.”

Leonardo Da Vinci.

Elements of a Mathematical Programming Model

- Decision Variables: The unknown quantities in a model that are assigned values when solving a problem.
  - Often denoted $x_1,\ldots,x_n$.

Elements of a Mathematical Programming Model

- Parameters
  - Objective Function Coefficients: Values that indicate the contribution to the objective function made by a one-unit increase in the corresponding decision variable(s).
  - Usually denoted as c_{i} for the decision variable x_i.

Elements of a Mathematical Programming Model

- Parameters
  - Constraint Coefficients: Values that indicate the contributions to the left-hand side of constraints made by a one-unit increase in the corresponding decision variable(s).
  - Usually denoted as a_{ij} for the decision variable x_i in constraint j.

Elements of a Mathematical Programming Model

- Parameters - Known values in the model
  - Right-Hand Side Values: Values used to express limitations placed on the left-hand side of constraints.
  - Usually denoted as b_j for constraint j.

Elements of a Mathematical Programming Model

- Constraints: Mathematical expressions that define limitations on the values of the decision variables as functions of the decision variables, constraint coefficients, and corresponding right-hand side value.
  - Often expressed as $g_j(a_{1j}x_1,\ldots,a_{nj}x_n|b_j)$
  - Indexed $j = 1,\ldots,m$

Components of a Mathematical Programming Model

- Feasible Region: Collection of potential values of the decision variables that simultaneously satisfy all *m* constraints.
- Objective Function: Mathematical expression that defines the goal of the model as a function of the decision variables and constraint coefficients.
  - Often expressed as $f(c_1x_1,\ldots,c_nx_n)$
  - Either maximized or minimized

Generic MP Model

- Scalar form:

  Optimize (max or min) $f(c_1x_1,\ldots,c_nx_n)$

  Subject to (st) $g_j(a_{1j}x_1,\ldots,a_{nj}x_n|b_j)$, $j = 1,\ldots,m$

- Matrix form:

  Optimize (max or min) $f(cx)$

  Subject to (st) $g(Ax|b)$

  where c is a 1xn vector

  x is a nx1 vector

  A is a mxn matrix

  b is a mx1 vector

Classes of MP Models

- Deterministic: The values of all parameters (a_is, b_js, c_is) are known with certainty and remain constant over the time horizon of the problem described by the model.

  This reality from the modeler perspective – this is how most mathematical programming problems are treated.

Classes of MP Models

- Stochastic: The value of at least one parameter parameters (a_is, b_js, c_is) is not known with certainty but can be represented by a known probability distribution that describes potential values of the parameter throughout the time horizon of the problem described by the model.

  This is reality from the modeler perspective if the modeler acknowledges the inherent uncertainty in the mathematical programming model.

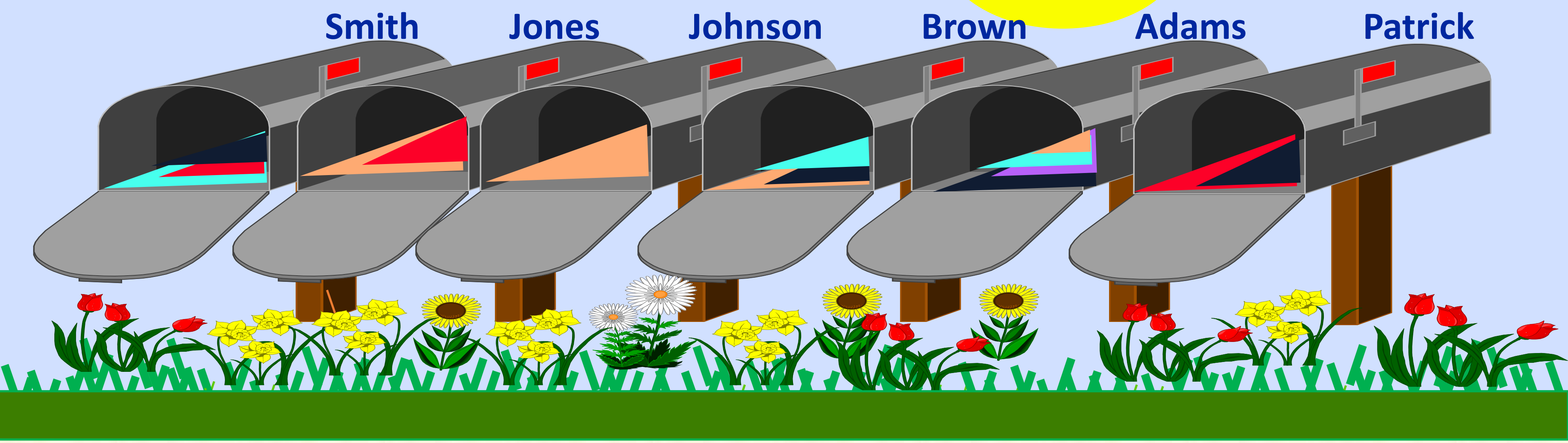
Classes of MP Models
A Broader Paradigm

- Sample-Based: The value of at least one parameter parameters (a_is, b_js, c_is) is not known with certainty and cannot be represented by a known probability distribution that describes potential values of the parameter throughout the time horizon of the problem described by the model, so is estimated using sample data.

  This is reality from a problem perspective (many MP models include parameters estimated with sample data but are treated deterministically)

THE MAGAZINE SUBSCRIPTION PROBLEM

N = 6 subscribers



M = 5 magazines

Cooking with Plums

Modern Calligraphy

Yogurt Digest

Amateur Acupuncture

Weasel Farming

population of *N* = 6

Components of the *Magazine Subscription Problem*

j = 1 (Cooking With Plums)

2 (Modern Calligraphy)

3 (Yogurt Digest)

4 (Amateur Acupuncture)

5 (Weasel Farming)

i = 1 (Smith), 2 (Jones), 3 (Johnson), 4 (Brown), 5 (Adams), 6 (Patrick)

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}, \boldsymbol{x}^{(1)'} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{x}^{(2)'} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{x}^{(3)'} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{x}^{(4)'} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

Feasible Region

$$\boldsymbol{x}^{(5)'} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{x}^{(6)'} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{x}^{(7)'} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{x}^{(8)'} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{x}^{(9)'} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{x}^{(10)'} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Mathematical Definition of the Magazine Subscription Problem

Suppose there are $\binom{M}{m}$ sets

$$\mathscr{S}_k = \left\{ i \mid \boldsymbol{a}_i \boldsymbol{x}_k' \geq 1 \right\}$$

where a_i is an M-dimensional vector ∍ $a_{ij} \in \{0,1\}$

x_k is an M-dimensional vector ∍ $x_{ik} \in \{0,1\}$

The optimal covering problem is

$$\max_k \left| \mathscr{S}_k \right|$$

Mathematical Definition of the Magazine Subscription Problem

Note that this problem can easily be augmented to allow for differential weights c_i to be applied to the vectors a_i.

$$\max_k \sum_{i \in \mathcal{S}_k} c_i$$

Full Enumeration for Magazine Subscription Problem

The coverages associated with the feasible space are

$$\mathscr{S}_1 = \{1,2,4,5,6\},\ \mathscr{S}_2 = \{1,4,5\}, \mathscr{S}_3 = \{1,2,4,5,6\},$$

$$\mathscr{S}_4 = \{1,2,3,4,5,6\}, \mathscr{S}_5 = \{1,2.3,4,5,6\}, \mathscr{S}_6 = \{2,3,4,5\},$$

$$\mathscr{S}_7 = \{1,4,5,6\}, \mathscr{S}_8 = \{1,2,5,6\}, \mathscr{S}_9 = \{4,5\}, \mathscr{S}_{10} = \{2,3,4,5\}$$

So the solutions associated with the feasible space are

$$|\mathscr{S}_1| = 5,\ |\mathscr{S}_2| = 3, |\mathscr{S}_3| = 5, |\mathscr{S}_4| = 6, |\mathscr{S}_5| = 6,$$

$$|\mathscr{S}_6| = 4, |\mathscr{S}_7| = 4, |\mathscr{S}_8| = 4, |\mathscr{S}_9| = 2, |\mathscr{S}_{10}| = 4$$

Magazine Subscription Problem
Optimal Solution

…and the optimal solution is to advertise in

a) Cooking With Plums & Amateur Acupuncture (x_1 & x_4)

or

b) Modern Calligraphy & Amateur Acupuncture (x_2 & x_4)

and achieve 100% coverage

$$\max_k |\mathscr{S}_k| = 6$$

The General Integer Programming Formulation

Maximize $\boldsymbol{c}'\boldsymbol{y}$

Subject to $\boldsymbol{A}\boldsymbol{x}' \geq \boldsymbol{y}$

$\mathbf{1}_m \boldsymbol{x}' = m$

$x_j, y_i \in \{0,1\} \; \forall \; i, j$

Magazine Subscription Problem
Integer Programming Formulation

Maximize $0y_4 + 0y_8 + 0y_9 + 0y_{11} + 0y_{14} + 0y_{30}$

Subject to

$$
\begin{aligned}
x_1 + x_2 + x_3 &\geq y_4 \text{ (Smith)} \\
x_2 + x_4 &\geq y_8 \text{ (Jones)} \\
x_4 &\geq y_9 \text{ (Johnson)} \\
x_1 + x_3 + x_4 &\geq y_{11} \text{ (Brown)} \\
x_1 + x_3 + x_4 + x_5 &\geq y_{14} \text{ (Adams)} \\
x_1 + x_2 &\geq y_{30} \text{ (Patrick)} \\
x_1 + x_2 + x_3 + x_4 + x_5 &= 2 \\
x_j, y_i \in \{0,1\} \ \forall\, i, j &
\end{aligned}
$$

Magazine Subscription Problem Optimal Solution

…and the optimal solution is to advertise in

a) Cooking With Plums & Amateur Acupuncture (x_1 & x_4)

or

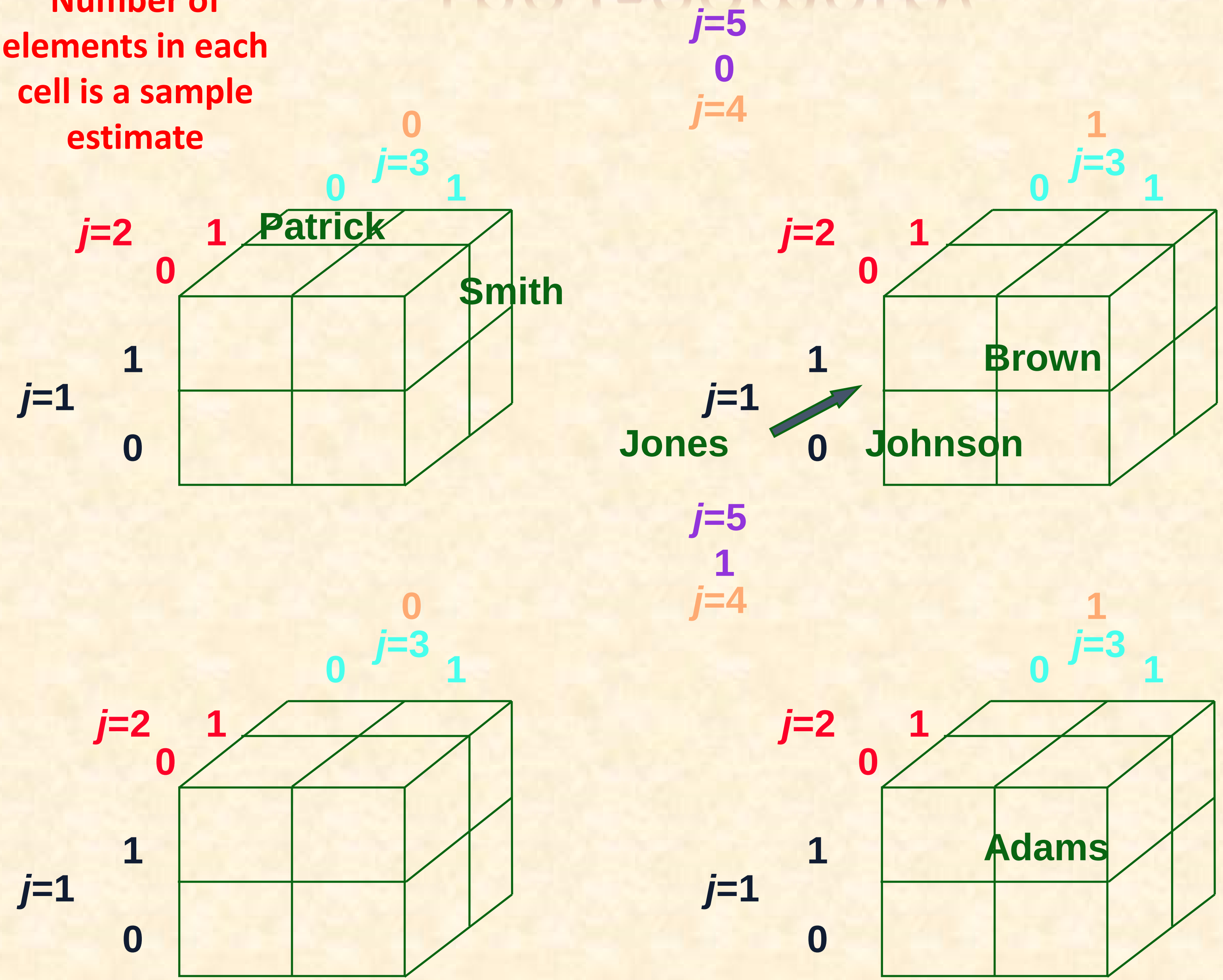
b) Modern Calligraphy & Amateur Acupuncture (x_2 & x_4)

and achieve 100% coverage

$$\max_k |\mathscr{S}_k| = 6$$

The Geometry

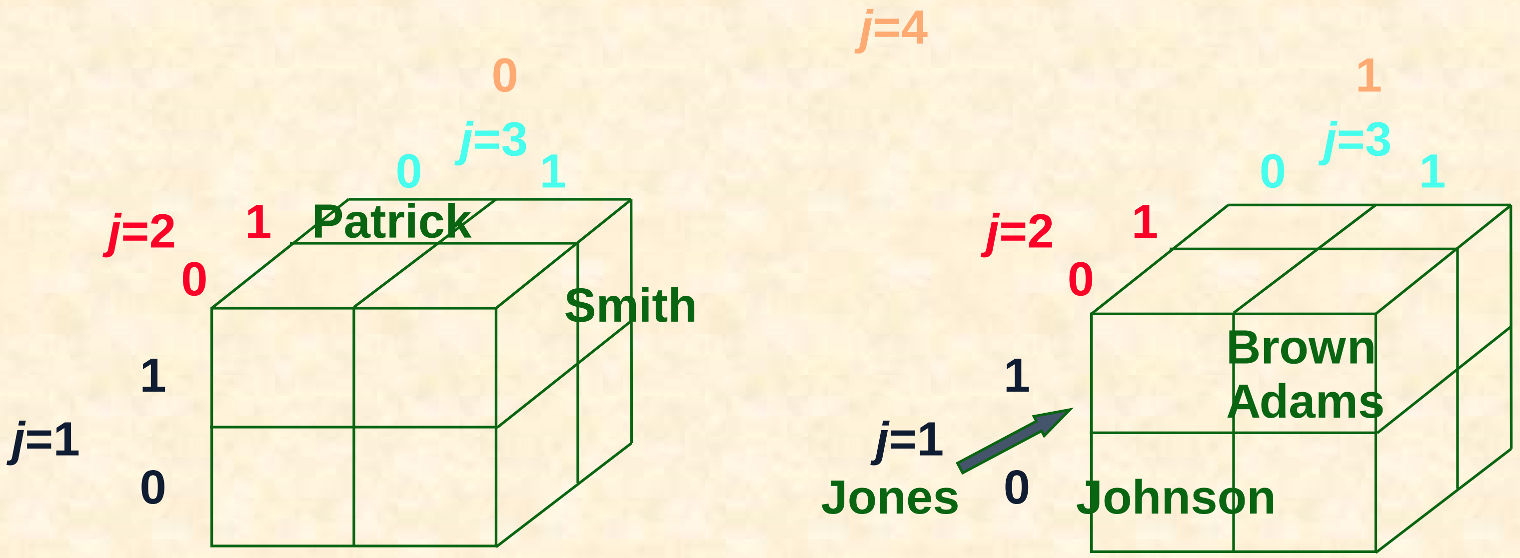
Number of elements in each cell is a sample estimate



The Geometry

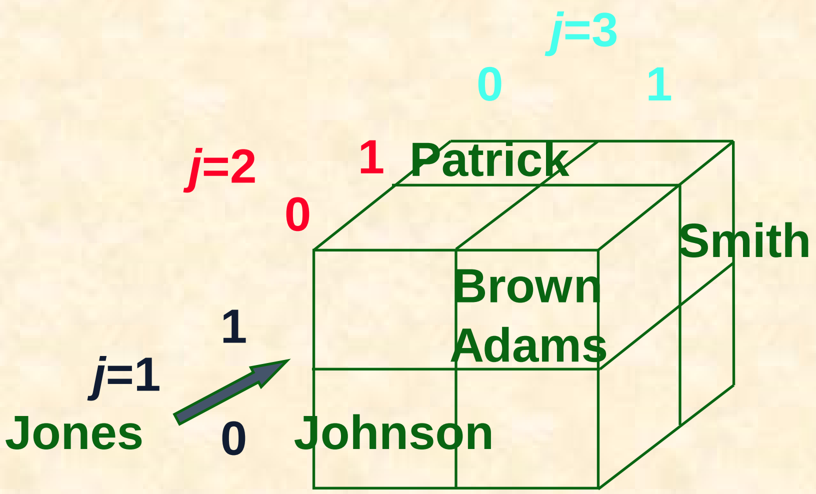
We wish to reduce the hypercube dimensions from *M* = 5 to *m* = 2 in the manner that minimizes the count in the ‘cell of non-coverage’

To evaluate coverage for *j* = 1 (*Cooking With Plums*) and *j* = 2 (*Modern Calligraphy*), we might first collapse (marginalize) on *j* = 5 (*Weasel Farming*)



The Geometry

We might then collapse on *j* = 4 (*Amateur Acupuncture*)



The Geometry

Finally, we collapse on j = 3 (*Yogurt Digest*)

		j=2: 0	*j*=2: 1
j=1	1	Brown Adams	Smith Patrick
	0	Johnson	Jones

Advertising in j = 1 (Cooking With Plums) and j = 2 (Modern Calligraphy) covers 83.33% (5/6) of the entities/subscribers.

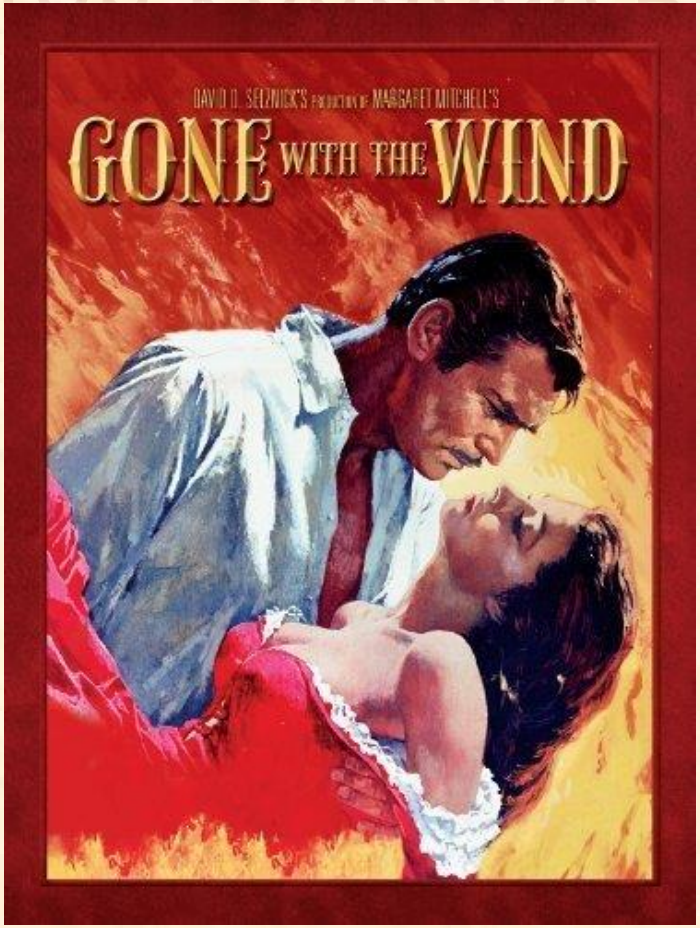
Magazine Subscription Problem Revisited

Real instances often include

- Hundreds of "magazines"
- Millions of "subscribers"

This is actually an instance of a class of combinatorial optimization problem referred to as *maximum* (or *optimal*) *covering*

Katie Scarlett O'Hara Hamilton Kennedy Butler?



Extending the Optimal Coverage Problem

Why was I intrigued by the possibility of working with a major studio to design the optimal plot for the sequel to *Gone With The Wind*?

- The problem of optimizing a movie plot was *very nontraditional/unique*
- The problem of optimizing a movie plot would present *many interesting modeling challenges*
- The notion of optimizing a movie plot *upset many purists* ☺
- Marketers *had not truly utilized constrained optimization/mathematical programming*

Optimal Product Design

Attributes & Attribute Levels?

Transmission:
- **Automatic**
- **Semi-Automatic**
- **Manual**

Roof:
- **Hardtop**
- **Moon Roof**
- **Convertible**

Color:
- **Red**
- **Blue**
- **Silver**

The Devaux Coupe

Notation for General IP Formulation

n = the number of respondents ($i = 1,\ldots,n$)

K = number of attributes ($k = 1,\ldots,K$)

L_k = number of levels for attribute k ($\ell = 1,\ldots,L_k$)

$\mu_{k\ell}^{(i)}$ = fractional measure of the value respondent i associates with level ℓ of attribute k when 'building' a prototype product

h_i = minimum utility to cover the i^{th} respondent

$x_{k\ell}$ = 1 if level ℓ of attribute k is chosen, 0 otherwise

y_i = 1 if respondent i is captured/satisfied, 0 otherwise

Generalized Share of Choices Problem IP Formulation

$$\text{Max} \sum_{i=1}^{n} c_i y_i$$

$$\text{st} \quad \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} u_{k\ell}^{(i)} x_{k\ell} \geq h_i y_i \qquad i = 1,2,\ldots,n$$

$$\sum_{\ell=1}^{L_k} x_{k\ell} = 1 \qquad k = 1,2,\ldots,K$$

$$x_{k\ell} \in \{0,1\} \qquad \ell = 1,2,\ldots,L_k, k = 1,2,\ldots,K$$

$$y_i \in \{0,1\} \qquad i = 1,2,\ldots,n$$

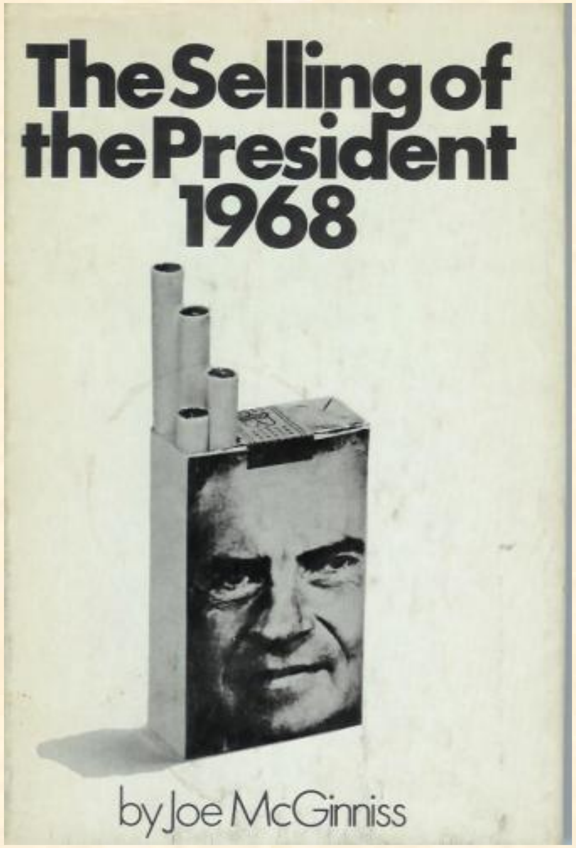
Generalized Share of Choices Problem IP Formulation

How do we meet the challenges of this problem when formulated through sampling?

- intelligent prepossessing - combine objects with identical ‘covering set membership profiles’
- intelligent algorithm development
  - branch and bound (and cut)
  - branching on the $x_{k\ell}$
  - logical pegging of the y_i at each branch
- Estimate solution and its variance and bias
  - Bootstrapping
  - Maintain collection of incumbent solutions

Optimal Political Platform Design

The inspiration for this research…



President Richard Milhaus Nixon

Optimal Political Platform Design

How can we apply the product optimization framework to finding the optimal political platform (a candidate's collection of positions on the salient issues)?

- *Issues* are analogous to *product attributes*
- *Positions* are analogous to *levels of product attributes*

Wow – maybe we can…

The Ramifications?

This leads to the following very interesting concerns:

- How likely have we identified the population optimal collection of attribute levels?
- How much variance is there in our estimate of the coverage achieved at optimality?
- Is there a bias in our estimate of the coverage achieved at optimality? If so, how large does this bias tend to be and in which direction does it tend to run?

Sample Based Optimization: Why Bother?

Understanding these characteristics enable us to

- produce a bias-adjusted estimate of the maximand or minimand
- estimate necessary sample size to achieve a level of precision in the estimated maximand or minimand
- build confidence intervals and test hypotheses about the estimated maximand or minimand
- understand the likelihood the sample-based solution is not the population optimal

Sample Based Optimization: Why Bother?

Understanding these characteristics enable us to

- estimate the probability the maximand or minimand will fall within some range of interest
- justify the adoption of a solution that is suboptimal for the sample data
- allow for development of Bayesian and predictive approaches that integrate prior information and/or beliefs

THE MORAL OF THESE EXAMPLES?

Analytics!

It is also important to note that here we see analysis of a big data problem through the integration of:

- Operations Research
- Statistics
- Probability
- Development of Algorithms
- Data Management

These research questions could not be addressed adequately if any one discipline was omitted

“The modeler must see what others only catch a glimpse of.”

Wrap Up
jjcochran@ua.edu

Wah wah wah! Kamsahamnida! Asante Sana! A houanu ka ka!
Mossi! Mə làbtə o!
ďakujem! Aitäh! Teşekkür ederim! Gracias!
Mèsi! dʑɛ̃'kujɛ̃!
Köszönöm! 谢谢 Mulțumesc! Arigato! Спасибо!
Vinaka! Thank You! Sag boluň!
děkuji! Dzięki! Obrigado!
Köszönöm Szépen! Köszönöm! Ilakasugotia!
Ke a leboha!
ຂອບໃຈ! धन्यवाद
Hvala vam! Ewe! Mauruuru!
ευχαριστώ Дякую!
Dhanyavaad! Спасибо! សូមអរគុណ
Hvala lijepa!
Danke! a dank aykh! Sanco!
გმადლობთ! Kiitos! Paldies!
Merci! ありがとう Ta kumbiiri!
Спасибо! agradeço!
dank je! Ngiyabonga! Abuimgang!
Ačiū! Дякую! شكرا لك Mulțumiri! Tänan!
Köszönöm! Kiwi Lingo!
Terima kasih!
Enkosi! Grazie! Благодарам
cảm ơn bạn! E se é! Takk fyrir!
дзякуй!