15.094J: Robust Modeling, Optimization, Computation

Lecture 18: Constructing Utilities using RO

Learning Preferences

- Kahneman and Tversky [1979] proposed prospect theory as a psychologically more accurate description of preferences compared to expected utility theory
  - Human behavior is inconsistent
  - People are loss averse
  - Gain or loss of an extra unit has less impact (risk averse)

- Our Goal: algorithmize prospect theory to ask questions and compute preferences in a tractable manner

Adaptive Questionnaires

- Items are represented by vectors of attributes

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

- Assume a linear utility function

$$u(\mathbf{x}) = \mathbf{u}'\mathbf{x}, \qquad \mathbf{u} \in \mathcal{U}^0 = [-1, 1]^n.$$

- If question is between $\mathbf{x}$ and $\mathbf{y}$, either

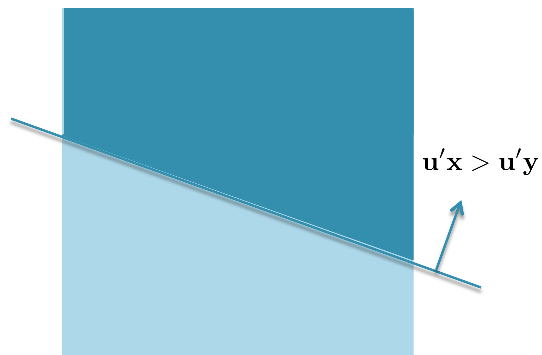
$$\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y} \quad \text{or} \quad \mathbf{u}'\mathbf{x} < \mathbf{u}'\mathbf{y}$$

Adaptive Questionnaires

- Initial feasible space $\mathcal{U}^0 = [-1, 1]^n$

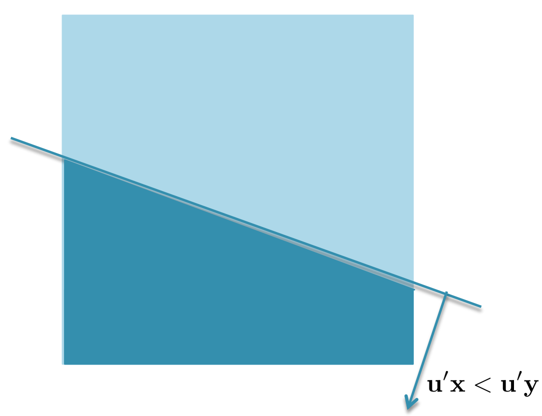
Adaptive Questionnaires

- Initial feasible space $\mathcal{U}^0 = [-1, 1]^n$
- Each question will result in a linear inequality
  - $\mathbf{x}$ or $\mathbf{y}$?
    - If $\mathbf{x}$, $\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}$



Adaptive Questionnaires

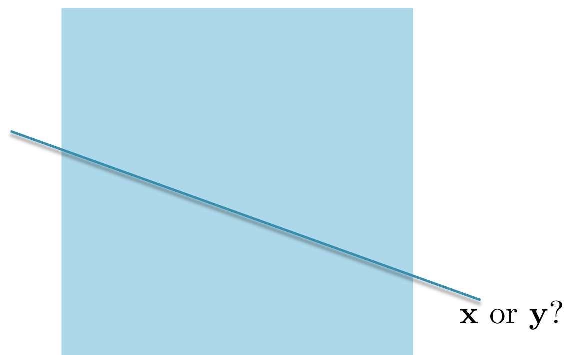
- Initial feasible space $\mathcal{U}^0 = [-1, 1]^n$
- Each question will result in a linear inequality
  - $\mathbf{x}$ or $\mathbf{y}$?
    - If $\mathbf{x}$, $\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}$
    - If $\mathbf{y}$, $\mathbf{u}'\mathbf{x} < \mathbf{u}'\mathbf{y}$



Adaptive Questionnaires

- Initial feasible space $\mathcal{U}^0 = [-1,1]^n$

- Each question will result in a linear inequality
  - $\mathbf{x}$ or $\mathbf{y}$?
    - If $\mathbf{x}$, $\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}$
    - If $\mathbf{y}$, $\mathbf{u}'\mathbf{x} < \mathbf{u}'\mathbf{y}$

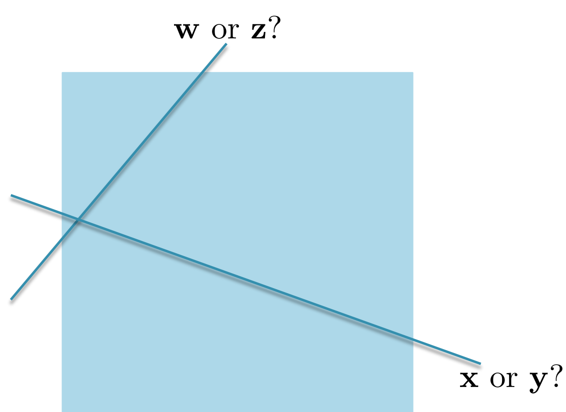


Adaptive Questionnaires

- Initial feasible space $\mathcal{U}^0 = [-1,1]^n$

- Each question will result in a linear inequality
  - $\mathbf{x}$ or $\mathbf{y}$?
    - If $\mathbf{x}$, $\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}$
    - If $\mathbf{y}$, $\mathbf{u}'\mathbf{x} < \mathbf{u}'\mathbf{y}$

  - $\mathbf{w}$ or $\mathbf{z}$?
    - If $\mathbf{w}$, $\mathbf{u}'\mathbf{w} < \mathbf{u}'\mathbf{z}$
    - If $\mathbf{z}$, $\mathbf{u}'\mathbf{w} > \mathbf{u}'\mathbf{z}$

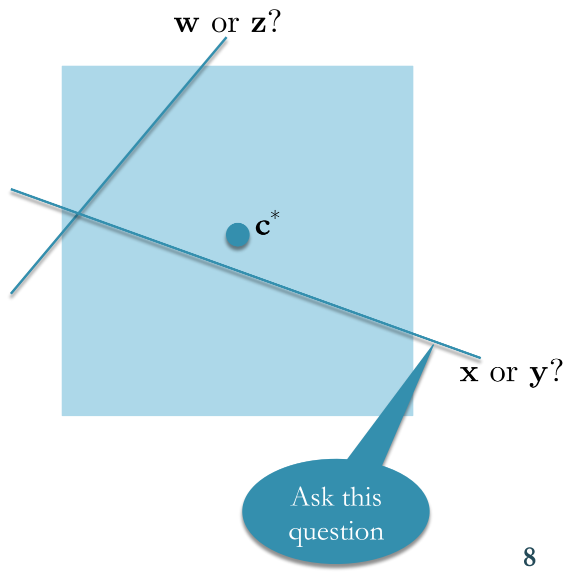


Adaptive Questionnaires

- Initial feasible space $\mathcal{U}^0 = [-1, 1]^n$

- Each question will result in a linear inequality
  - $\mathbf{x}$ or $\mathbf{y}$?
    - If $\mathbf{x}$, $\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}$
    - If $\mathbf{y}$, $\mathbf{u}'\mathbf{x} < \mathbf{u}'\mathbf{y}$
  - $\mathbf{w}$ or $\mathbf{z}$?
    - If $\mathbf{w}$, $\mathbf{u}'\mathbf{w} < \mathbf{u}'\mathbf{z}$
    - If $\mathbf{z}$, $\mathbf{u}'\mathbf{w} > \mathbf{u}'\mathbf{z}$

- Pick the question that cuts $\mathcal{U}^0$ closest to the analytic center $\mathbf{c}^*$



Adaptive Questionnaires

- Add inequality with small $\epsilon > 0$

$$\mathbf{u}'\mathbf{x} \geq \mathbf{u}'\mathbf{y} + \epsilon \qquad \text{or} \qquad \mathbf{u}'\mathbf{x} \leq \mathbf{u}'\mathbf{y} - \epsilon$$

- Compute analytic center of new polytope

- Select next question closest to analytic center

- Repeat until question limit or feasible space can not be reduced

Adaptive Questionnaires

- Introduced in marketing [Toubia et al. 2003, 2004]
- Results often suffer from response errors
  - Incorrect responses influence later questions
- We use integer and robust optimization

Self-Correcting Mechanism

- Accounts for inconsistencies and response errors
- Binary variable ϕ_i for each question i $(\mathbf{x}^i > \mathbf{y}^i)$

$$\mathbf{u}'(\mathbf{x}^i - \mathbf{y}^i) + (n + \epsilon)\phi_i \geq \epsilon$$
$$\mathbf{u}'(\mathbf{x}^i - \mathbf{y}^i) + (n + \epsilon)\phi_i \leq n$$

  - If $\phi_i = 0$, constraints are consistent with response
  - If $\phi_i = 1$, a response error or inconsistency is assumed

Compute the Analytic Center

- For the next question, compute the analytic center of

$$\begin{aligned}\mathcal{U}^k = \{\mathbf{u} \in \mathbb{R}^n \mid\ & \mathbf{u}'(\mathbf{x}^i - \mathbf{y}^i) + (\epsilon + n)\phi_i \geq \epsilon, \\ & \mathbf{u}'(\mathbf{x}^i - \mathbf{y}^i) + (\epsilon + n)\phi_i \leq n, \\ & -1 \leq u_j \leq 1, \\ & \sum_{i=1}^{k} \phi_k \leq \gamma k \\ & \phi_i \in \{0, 1\} \quad \}.\end{aligned}$$

Allow a small fraction of the constraints to "flip"

- After k questions, feasible utilities are given by the set $\mathcal{U}^k \subseteq \mathcal{U}^0$
  - $\mathcal{U}^k$ is the projection of a discrete set – union of polyhedra

Loss Averse Preferences

- Previously, the analytic center was taken to be the user's utility vector

- Instead, we solve

$$\max_{\mathbf{x} \in \mathbf{X}} \ \min_{\mathbf{u} \in \mathcal{U}^k} \mathbf{u}'\mathbf{x}$$

where the uncertainty set $\mathcal{U}^k$ is the outcome of a dynamic questionnaire

Robust Optimization with CVaR

- We want to balance robustness and optimality
- Maximize Conditional Value at Risk (CVaR)
  - Definition: CVaR at level α is the expected value of the worst $\alpha\%$ of the utilities in $\mathcal{U}^k$
  - Example:
    - Utilities of 5 different items: (-0.75, -0.25, 0.1, 0.3, 0.8)
      - CVaR(40%) = -0.50
    - Compare to VaR (α-quantile) and Standard Deviation
      - VaR(40%) = -0.25 Doesn't account for -0.75!
      - StdDev = 0.62 Increases for gains and loses!
- CVaR captures the **amount of losses**

Robust Optimization with CVaR

- Challenge: $\mathcal{U}^k$ is the union of polyhedra
  - Fix ϕ to the optimal values in the final computation of the analytic center

$$\Longrightarrow \mathcal{U}^k \text{ becomes a polyhedron}$$

- Challenge: CVaR of $\mathcal{U}^k$ is defined by an integral
  - Approximate $\mathcal{U}^k$ with random sampling
  - We use "Hit-and-Run," starting from $\mathbf{c}^*$ to sample N utility vectors

$$\{\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^N\} \in \mathcal{U}^k$$

  - Polynomial time algorithm, fast in practice [Lovász, Vempala 2006]

Robust Optimization with CVaR

- Given $\{\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^N\} \in \mathcal{U}^k$, solve

$$\begin{aligned} \max_{\mathbf{x}\in X} \quad \min_{\mathbf{y}} \quad & \frac{1}{\alpha N}\sum_{j=1}^{N}(\mathbf{u}_j'\mathbf{x})y_j \\ \text{s.t.} \quad & \sum_{j=1}^{N} y_j = \alpha N \\ & 0 \le y_j \le 1 \end{aligned}$$

- Robust approach: $\alpha = \frac{1}{N}, \quad N \to \infty$

Empirical Evidence

- Experiment
  - Randomly select a "true" utility vector $\mathbf{u}^* \in \mathcal{U}^0$
  - Answer questions according to $\mathbf{u}^*$ with normally distributed noise $\zeta \sim N(0, \sigma)$

$$(\mathbf{u}^*)'(\mathbf{x}^i - \mathbf{y}^i) + \zeta \ge 0 \Longrightarrow \mathbf{x}^i > \mathbf{y}^i$$
$$(\mathbf{u}^*)'(\mathbf{x}^i - \mathbf{y}^i) + \zeta < 0 \Longrightarrow \mathbf{y}^i > \mathbf{x}^i$$

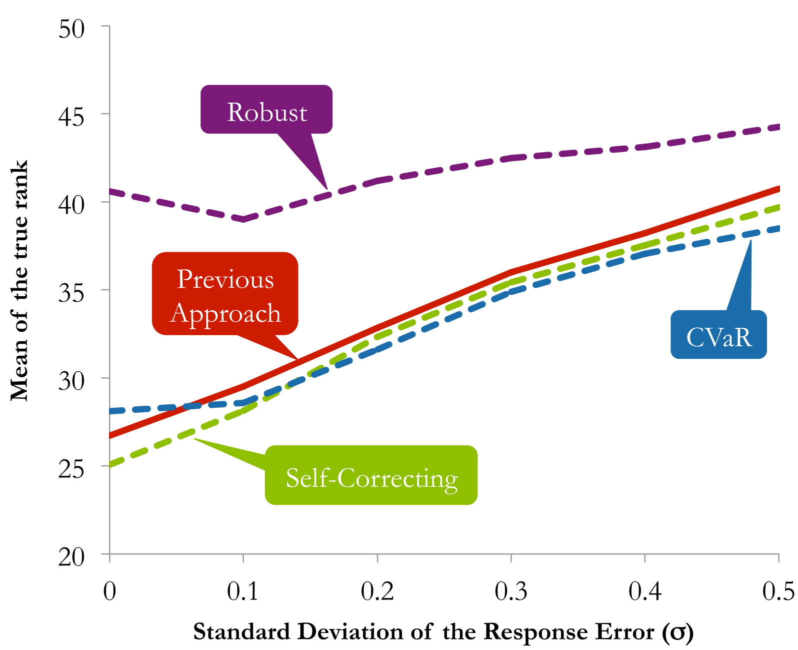
Empirical Evidence

- Rank all items according to $\mathbf{u}^*$ (from 1 to 102)

Item	1	2	3	4	5	6	7	8	9	10	…
Rank	32	3	10	55	2	8	15	4	27	80	…

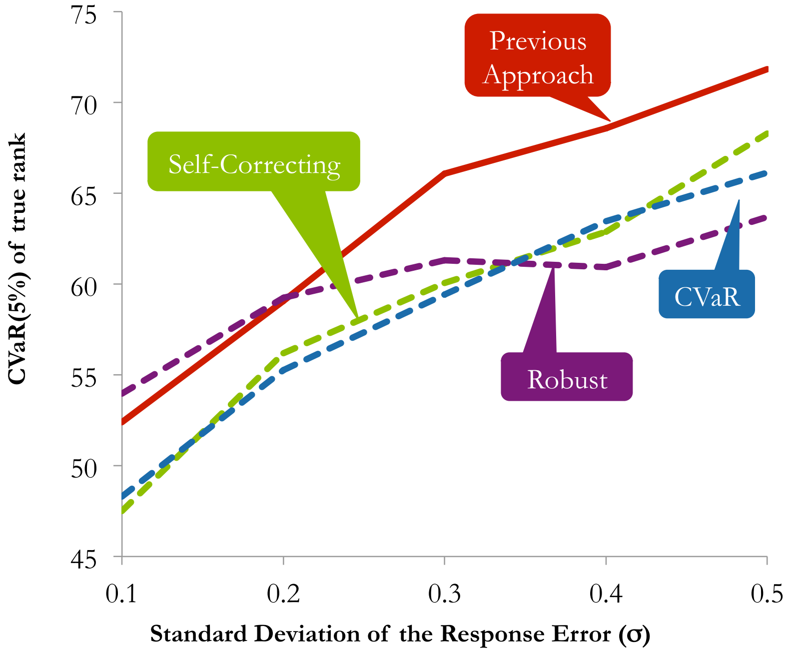
- Solve for the best 5 items according to our algorithm
  - We assume the user will be given choices
  - Example: Items 2, 5, 7, 8, 9
- Compute the average true rank of the solutions we found
  - Example: 10.2
  - Smaller is better

Average Value of the True Ranks



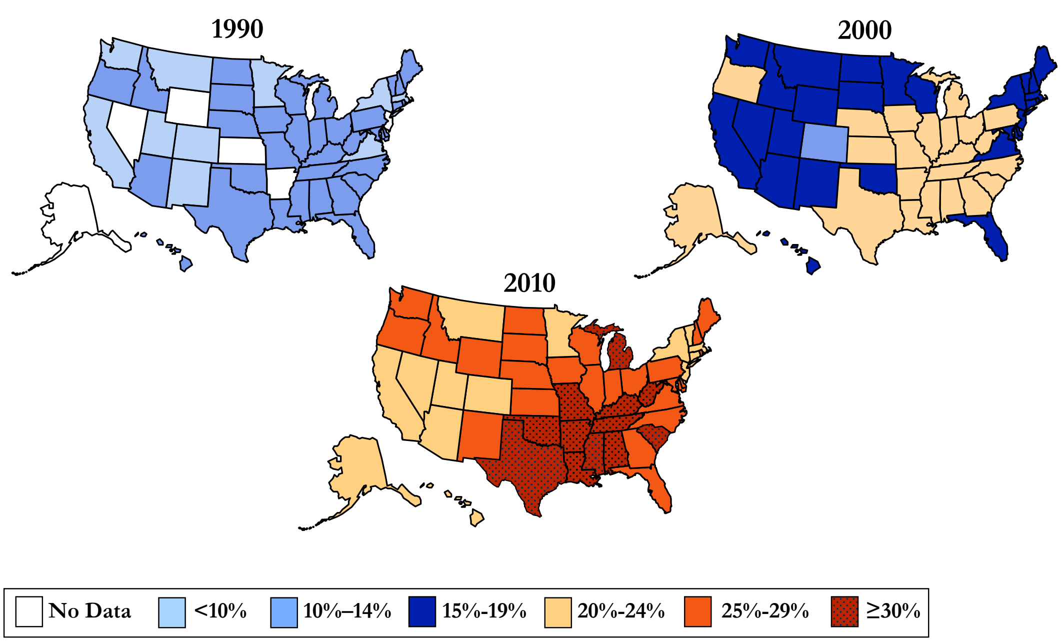
- Average over 10 users
- 15 questions
- $\gamma = 0.1$ for all methods
- $\alpha = 0.1$ for CVaR

CVaR(5%) of the True Ranks



- Average over 10 users
- 15 questions
- $\gamma = 0.1$ for all methods
- $\alpha = 0.1$ for CVaR

Obesity Trends Among U.S. Adults



Source: Behavioral Risk Factor Surveillance System, Centers for Disease Control and Prevention

Obesity and Diabetes in the US

- The incidence of type II diabetes is increasing with obesity
  - 25.8 million people in the US (8.3% of the population)
  - 35% of adults have prediabetes, 50% of adults 65+
- Leading causes of heart disease, stroke, blindness, amputation and kidney failure
- On average,
  - Obesity costs $1,429 more per person in medical costs each year
  - Diabetes costs more than twice as much per person

Diabetes and Dietary Adherence

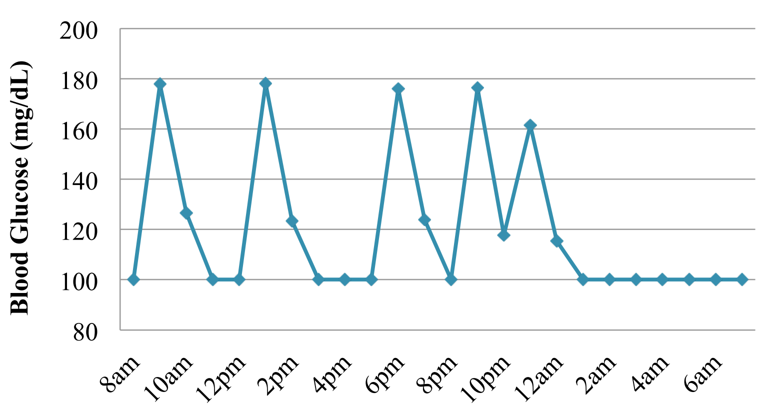
- "Patients with good diabetes self-care behaviors can attain excellent glycemic control…patients with diabetes are especially prone to substantial regimen adherence problems"
  - American Diabetes Association
- "There was substantial agreement by health professionals and patients alike that diet and diet-related issues constituted the most difficult problem faced by persons with diabetes and by health professionals caring for those persons."
  - Michigan Diabetes Research and Training Center
- "One way to improve dietary adherence rates in clinical practice may be to use a broad spectrum of diet options, to better match individual patient food preferences"
  - Dansinger et al., JAMA 2005

Our Approach

- We have developed a **personalized**, **comprehensive**, and **dynamic** system for diabetes management
- Learn food preferences to improve diet adherence
  - Self-Correcting Adaptive Questionnaires
  - A Robust CVaR Approach
- Model blood glucose dynamics
  - People respond differently to different foods
- Propose a daily food and exercise plan

Blood Glucose Dynamics

- Blood glucose level is measured in milligrams per decilitre (mg/dL)
- Blood glucose levels follow a trajectory (BG curve)

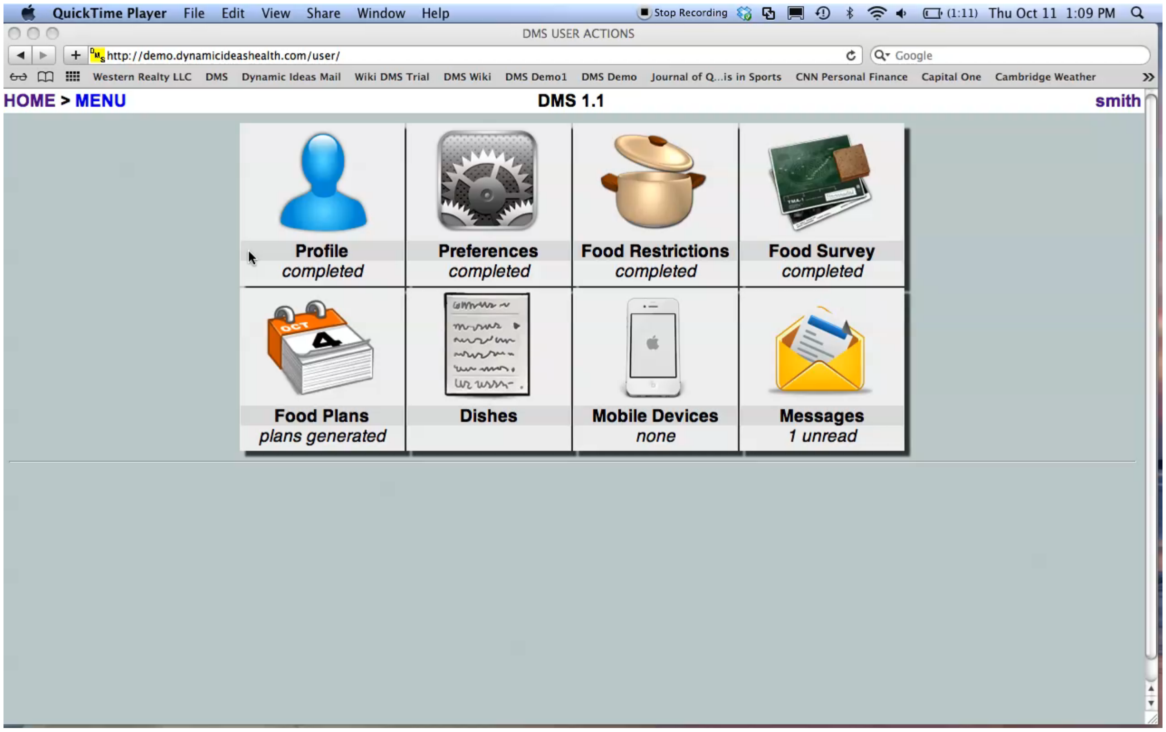


Blood Glucose Modeling

- We model the blood glucose levels as a function of
  - Fasting level
  - Food consumed
    - Classify foods into categories using the glycemic index
    - Measures the effects of carbohydrates on blood glucose levels
  - Exercise performed
- BG measurements can be used to learn the BG curve
  - Regression to update
  - Robust optimization to account for error

Plan Generation with MIO

- Maximize preferences
  - Penalties for high blood glucose levels, nutritional violations
- Bounds on max/min blood glucose levels
- Nutritional requirements
  - Calories, carbs, protein, fat, etc.
- Food group requirements
  - Fruits, vegetables, dairy, meat, starch
- "Appeal" constraints
  - Variety, timing, balanced meals



Summary

- Learn preferences consistent with human behavior
  - Self-Correcting Mechanism
  - A Robust CVaR approach

- An overall system to improve diabetes and diet management