On a Class of Uncertainty Sets for Distributionally Robust Optimization

Erick Delage

Assistant professor

Department of Management Sciences

HEC Montréal

Joint work with Y. Ye, S. Arroyo, and J. Carlsson.

Workshop on Robust Optimization
May 2011

Outline

Outline

Stochastic Programming Approach

Let's consider a stochastic programming problem:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \mathbb{E}\left[u(\mathsf{h}(\mathbf{x},\boldsymbol{\xi}))\right]$$

where $\mathbf{x}$ = decisions and $\boldsymbol{\xi}$ = uncertain parameters.

Stochastic Programming Approach

Let's consider a stochastic programming problem:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \mathbb{E}\left[u(\mathrm{h}(\mathbf{x},\boldsymbol{\xi}))\right]$$

where $\mathbf{x}$ = decisions and $\boldsymbol{\xi}$ = uncertain parameters.

Here, we assume :

- We know the distribution of the random vector $\boldsymbol{\xi}$
- The profit function $\mathrm{h}(\mathbf{x},\boldsymbol{\xi})$ is concave in $\mathbf{x}$ and convex in $\boldsymbol{\xi}$
- We have a piecewise linear utility function that characterizes investor's risk aversion

$$u(y) = \min_{k\in\{1,2,\dots,K\}} \alpha_k y + \beta_k$$

Difficulty of developing a probabilistic model

Developing an accurate probabilistic model requires heavy engineering efforts:

- Need to collect enough observations
- Need to consult with experts of the field of practice
- Need to make simplifying assumptions

Difficulty of developing a probabilistic model

Developing an accurate probabilistic model requires heavy engineering efforts:

- Need to collect enough observations
- Need to consult with experts of the field of practice
- Need to make simplifying assumptions

Yet, there are inherent pitfalls in the process:

- Expecting that a scenario might occur does not determine its probability of occurring
- Unexpected event (e.g., economic crisis) might occur
- The future might actually not behave like the past

Limits of Expected Utility: Ellsberg Paradox

Consider an urn with 30 blue balls and 60 other balls that are either red or yellow (you don't know how many are red or yellow).

Limits of Expected Utility: Ellsberg Paradox

Consider an urn with 30 blue balls and 60 other balls that are either red or yellow (you don't know how many are red or yellow).

Experiment 1: Choose among the following two gambles

- Gamble A: If you draw a blue ball, then you win 100$
- Gamble B: If you draw a red ball, then you win 100$

Limits of Expected Utility: Ellsberg Paradox

Consider an urn with 30 blue balls and 60 other balls that are either red or yellow (you don't know how many are red or yellow).

Experiment 1: Choose among the following two gambles

- Gamble A: If you draw a blue ball, then you win 100$
- Gamble B: If you draw a red ball, then you win 100$

Experiment 2: Choose among the following two gambles

- Gamble C: If you draw blue or yellow ball, then you win 100$
- Gamble D: If you draw red or yellow ball, then you win 100$

Limits of Expected Utility: Ellsberg Paradox

Consider an urn with 30 blue balls and 60 other balls that are either red or yellow (you don't know how many are red or yellow).

Experiment 1: Choose among the following two gambles

- Gamble A: If you draw a blue ball, then you win 100$
- Gamble B: If you draw a red ball, then you win 100$

Experiment 2: Choose among the following two gambles

- Gamble C: If you draw blue or yellow ball, then you win 100$
- Gamble D: If you draw red or yellow ball, then you win 100$

If you clearly prefer Gamble A & D, then you cannot be thinking in terms of expected utility.

The Distributionally Robust Formulation

- We consider that the choice of F is ambiguous
- Use available information to define $\mathcal{D}$, such that $F \in \mathcal{D}$
- Distributionally Robust Optimization values a decision using the lowest performing distribution

$$(DRSP) \qquad \underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \; \inf_{F\in\mathcal{D}} \; \mathbb{E}_F[u(\mathsf{h}(\mathbf{x},\boldsymbol{\xi}))]$$

The Distributionally Robust Formulation

- We consider that the choice of F is ambiguous
- Use available information to define $\mathcal{D}$, such that $F \in \mathcal{D}$
- Distributionally Robust Optimization values a decision using the lowest performing distribution

$$(DRSP) \qquad \underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \; \inf_{F\in\mathcal{D}} \mathbb{E}_F[u(\mathrm{h}(\mathbf{x},\boldsymbol{\xi}))]$$

  - Introduced by H. Scarf in 1958
  - Recently, we found ways of solving some DRSP's efficiently [Popescu (2007), Bertsimas et al., Natarajan et al., Delage et al. (2010)]
  - Growing interest for promoting performance differently depending on F [Föllmer et al. (2002), Li et al. (2011)]

Outline

Definition of $\mathcal{D}(\psi)$

Let $\mathcal{D}(\psi)$ be a set of the form

$$\mathcal{D}(\psi) = \left\{ F \in \mathcal{M} \,\middle|\, \begin{array}{l} \mathbb{P}_F(\boldsymbol{\xi} \in \mathcal{S}) = 1 \\ \mathbf{z}^{\mathsf{T}} \left(\mathbb{E}_F[\psi(\boldsymbol{\xi})] - \mathbf{b} \right) \leq 0 \ , \ \forall \mathbf{z} \in \mathcal{K} \end{array} \right\}$$

such that:

1. The set $\mathcal{S}$ is closed, convex, and bounded
2. The set $\mathcal{K} \subseteq \mathbb{R}^p$ is a convex cone
3. $\forall \mathbf{z} \in \mathcal{K}$, the function $g(\mathbf{z}, \boldsymbol{\xi}) := \mathbf{z}^{\mathsf{T}} \psi(\boldsymbol{\xi})$ is convex in $\boldsymbol{\xi}$
4. Both $\mathcal{S}$ and $\mathcal{K}$ have "nice" representations

$$\mathcal{S} = \{\boldsymbol{\xi} \in \mathbb{R}^m | g_i^S(\boldsymbol{\xi}) \leq 0 \, \forall \, i\}$$

$$\mathcal{K} = \{\mathbf{z} \in \mathbb{R}^p | g_j^K(\mathbf{z}) \leq 0 \, \forall \, j\}$$

$\mathcal{D}(\psi)$ imposes that F be Concentrated

Lemma

Given any random vector ζ such that $F_\zeta \in \mathcal{D}(\psi)$, for all $0 \leq \theta \leq 1$ the distribution of $\zeta' := \theta(\zeta - \mathbb{E}[\zeta]) + \mathbb{E}[\zeta]$ also has a distribution in $\mathcal{D}(\psi)$.

$\mathcal{D}(\psi)$ imposes that F be Concentrated

Lemma

Given any random vector ζ such that $F_\zeta \in \mathcal{D}(\psi)$, for all $0 \leq \theta \leq 1$ the distribution of $\zeta' := \theta(\zeta - \mathbb{E}[\zeta]) + \mathbb{E}[\zeta]$ also has a distribution in $\mathcal{D}(\psi)$.

Proof: (let $\boldsymbol{\mu} := \mathbb{E}[\zeta]$)

1. Since $\mathcal{S}$ is convex and $\mathbb{P}(\zeta \in \mathcal{S}) = 1$, we have that $\mathbb{E}[\zeta] \in \mathcal{S}$, hence
$$\mathbb{P}(\zeta' \in \mathcal{S}) = \mathbb{P}(\theta\boldsymbol{\zeta} + (1-\theta)\boldsymbol{\mu} \in \mathcal{S}) = 1$$
2. By Jensen's inequality for any $\mathbf{z} \in \mathcal{K}$, we have that
$$\begin{aligned}
\mathbb{E}[\mathbf{z}^T\psi(\theta\boldsymbol{\zeta} + (1-\theta)\boldsymbol{\mu})] &\leq \mathbb{E}[\theta\mathbf{z}^T\psi(\zeta) + (1-\theta)\mathbf{z}^T\psi(\boldsymbol{\mu})] \\
&= \theta\mathbb{E}[\mathbf{z}^T\psi(\zeta)] + (1-\theta)\mathbf{z}^T\psi(\boldsymbol{\mu}) \\
&\leq \mathbb{E}[\mathbf{z}^T\psi(\zeta)] \leq \mathbf{z}^T\mathbf{b}
\end{aligned}$$

Examples

- Upper bounds on partial moments : $\mathbb{E}[\max(0, \boldsymbol{\xi}_i - \hat{\mu}_i)^2] \leq b_i$ since equivalent to

$$z(\mathbb{E}[\max(0, \boldsymbol{\xi}_i - \hat{\mu}_i)^2] - b_i) \leq 0, \ \forall \, z \geq 0$$

Examples

- Upper bounds on partial moments : $\mathbb{E}[\max(0, \boldsymbol{\xi}_i - \hat{\mu}_i)^2] \leq b_i$ since equivalent to

$$z(\mathbb{E}[\max(0, \boldsymbol{\xi}_i - \hat{\mu}_i)^2] - b_i) \leq 0, \ \forall z \geq 0$$

- Confidence ellipsoid for mean :

$$(\hat{\boldsymbol{\mu}} - \mathbb{E}[\boldsymbol{\xi}])^T \hat{\Sigma}^{-1} (\hat{\boldsymbol{\mu}} - \mathbb{E}[\boldsymbol{\xi}]) \leq \gamma$$

since equivalent to:

$$\mathbb{E}\left[\mathbf{Z} \bullet \begin{bmatrix} \hat{\Sigma} & (\hat{\boldsymbol{\mu}} - \boldsymbol{\xi}) \\ (\hat{\boldsymbol{\mu}} - \boldsymbol{\xi})^T & \gamma \end{bmatrix}\right] \geq 0, \ \forall \mathbf{Z} \succeq 0$$

Examples

- Upper bounds on partial moments : $\mathbb{E}[\max(0, \boldsymbol{\xi}_i - \hat{\mu}_i)^2] \leq b_i$ since equivalent to

$$z(\mathbb{E}[\max(0, \boldsymbol{\xi}_i - \hat{\mu}_i)^2] - b_i) \leq 0, \ \forall z \geq 0$$

- Confidence ellipsoid for mean :

$$(\hat{\boldsymbol{\mu}} - \mathbb{E}[\boldsymbol{\xi}])^T \hat{\Sigma}^{-1} (\hat{\boldsymbol{\mu}} - \mathbb{E}[\boldsymbol{\xi}]) \leq \gamma$$

since equivalent to:

$$\mathbb{E}\left[\mathbf{Z} \bullet \begin{bmatrix} \hat{\Sigma} & (\hat{\boldsymbol{\mu}} - \boldsymbol{\xi}) \\ (\hat{\boldsymbol{\mu}} - \boldsymbol{\xi})^T & \gamma \end{bmatrix}\right] \geq 0, \ \forall \mathbf{Z} \succeq 0$$

- Upper bound on second order moment matrix

$$\mathbb{E}[(\boldsymbol{\xi} - \hat{\boldsymbol{\mu}})(\boldsymbol{\xi} - \hat{\boldsymbol{\mu}})^T] \preceq \gamma \hat{\Sigma}$$

since equivalent to:

$$\mathbb{E}[\mathbf{Z} \bullet ((\boldsymbol{\xi} - \hat{\boldsymbol{\mu}})(\boldsymbol{\xi} - \hat{\boldsymbol{\mu}})^T - \gamma \hat{\Sigma})] \leq 0, \ \forall \mathbf{Z} \succeq 0$$

Resolving Distributional Set from Data

- Question:
  - We have in hand i.i.d. samples $\{\boldsymbol{\xi}_i\}_{i=1}^M$
  - We know that $\mathbb{P}(\boldsymbol{\xi} \in \mathcal{S}) = 1$ and $\mathcal{S} \subseteq \mathcal{B}(\mathbf{0}, R)$
  - We can estimate the mean and covariance matrix:

$$\hat{\boldsymbol{\mu}} = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{\xi}_i \qquad \hat{\Sigma} = \frac{1}{M}\sum_{i=1}^{M}(\boldsymbol{\xi}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{\xi}_i - \hat{\boldsymbol{\mu}})^{\mathsf{T}}$$

  - What do we know about the distribution behind these samples?

Resolving Distributional Set from Data

- Question:
  - We have in hand i.i.d. samples $\{\boldsymbol{\xi}_i\}_{i=1}^{M}$
  - We know that $\mathbb{P}(\boldsymbol{\xi} \in \mathcal{S}) = 1$ and $\mathcal{S} \subseteq \mathcal{B}(\mathbf{0}, R)$
  - We can estimate the mean and covariance matrix:

$$\hat{\boldsymbol{\mu}} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{\xi}_i \qquad \hat{\Sigma} = \frac{1}{M}\sum_{i=1}^{M} (\boldsymbol{\xi}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{\xi}_i - \hat{\boldsymbol{\mu}})^{\mathsf{T}}$$

  - What do we know about the distribution behind these samples?

- Answer:

$$\mathcal{D}_2(\gamma) = \left\{ F \left| \begin{array}{l} \mathbb{P}(\boldsymbol{\xi} \in \mathcal{S}) = 1 \\ \|\mathbb{E}\left[\boldsymbol{\xi}\right] - \hat{\mu}\|^2_{\hat{\Sigma}^{-1/2}} \leq \gamma_1 \\ \mathbb{E}\left[(\boldsymbol{\xi} - \hat{\boldsymbol{\mu}})(\boldsymbol{\xi} - \hat{\boldsymbol{\mu}})^{\mathsf{T}}\right] \preceq (1 + \gamma_2)\hat{\Sigma} \end{array} \right. \right\}$$

  - With prob. $> 1 - \delta$ the distribution is contained in $\mathcal{D}_2(\gamma)$ for some $\gamma_1 = O\left(\frac{R^2}{M}\log(1/\delta)\right)$ and $\gamma_2 = O\left(\frac{R^2}{\sqrt{M}}\sqrt{\log(1/\delta)}\right)$.

Verifying whether $\mathcal{D}(\psi)$ is Empty or Not

- One finds a distribution that lies in the relative interior of $\mathcal{D}(\psi)$ by verifying that $t^* > 0$ for the problem:

$$\begin{aligned} \max_{F\in\mathcal{M},t\in\mathbb{R}}. \quad & t \\ \text{s.t.} \quad & \mathbf{z}^{\mathsf{T}}(\mathbb{E}_F[\psi(\boldsymbol{\xi})] - \mathbf{b} + t\mathbf{z}_0^*) \leq 0 \,, \; \forall\, \mathbf{z} \in \mathcal{K} \\ & \mathbb{P}_F(\boldsymbol{\xi} \in \mathcal{S}) = 1 \,, \end{aligned}$$

  where $\mathbf{z}_0^*$ is any vector lying in the strict interior of $\mathcal{K}^*$.

-

Verifying whether $\mathcal{D}(\psi)$ is Empty or Not

- One finds a distribution that lies in the relative interior of $\mathcal{D}(\psi)$ by verifying that $t^* > 0$ for the problem:

$$\max_{\boldsymbol{\mu}\in\mathcal{S},t\in\mathbb{R}}. \quad t$$
$$\text{s.t.} \quad \mathbf{z}^{\mathsf{T}}(\psi(\boldsymbol{\mu}) - \mathbf{b} + t\mathbf{z}_0^*) \leq 0 \ , \ \forall\, \mathbf{z} \in \mathcal{K} \ ,$$

where $\mathbf{z}_0^*$ is any vector lying in the strict interior of $\mathcal{K}^*$.
  - Remember that $\psi(\mathbb{E}_F[\boldsymbol{\xi}]) \leq \mathbb{E}_F[\psi(\boldsymbol{\xi})] \ , \forall\, F \in \mathcal{M}$.

Verifying whether $\mathcal{D}(\psi)$ is Empty or Not

- One finds a distribution that lies in the relative interior of $\mathcal{D}(\psi)$ by verifying that $t^* > 0$ for the problem:

$$\begin{aligned} \max_{\boldsymbol{\mu}\in\mathcal{S}, t\in\mathbb{R}}. \quad & t \\ \text{s.t.} \quad & \mathbf{z}^{\mathsf{T}}(\psi(\boldsymbol{\mu}) - \mathbf{b} + t\mathbf{z}_0^*) \leq 0 \, , \; \forall\, \mathbf{z} \in \mathcal{K} \, , \end{aligned}$$

  where $\mathbf{z}_0^*$ is any vector lying in the strict interior of $\mathcal{K}^*$.
  - Remember that $\psi(\mathbb{E}_F[\boldsymbol{\xi}]) \leq \mathbb{E}_F[\psi(\boldsymbol{\xi})] \, , \forall\, F \in \mathcal{M}$.

- This problem verifies that $\exists\, \boldsymbol{\mu} \in \mathcal{S}$ such that $\mathbf{b} - \psi(\boldsymbol{\mu})$ is in the strict interior of $\mathcal{K}^*$, i.e. a “simple” finite dimensional constraint qualification verification.

Solving the DRSP problem

- Given that $\exists F \in \mathcal{D}(\psi)$ that strictly satisfies all moment constraints, then one can apply duality and solve the equivalent problem:

$$\begin{array}{lll} \underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} & \underset{F\in\mathcal{M}}{\inf} & \mathbb{E}_F[u(h(\mathbf{x},\boldsymbol{\xi}))] \\ & \text{s.t.} & \mathbb{P}_F(\boldsymbol{\xi}\in\mathcal{S}) = 1 \\ & & \mathbf{z}^\mathsf{T}(\mathbb{E}_F[\psi(\boldsymbol{\xi})] - \mathbf{b}) \leq 0 \;,\; \forall\, \mathbf{z}\in\mathcal{K} \end{array}$$

-

Solving the DRSP problem

- Given that $\exists F \in \mathcal{D}(\psi)$ that strictly satisfies all moment constraints, then one can apply duality and solve the equivalent problem:

$$\begin{aligned} \underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \max_{r,\mathbf{z}} \quad & r - \mathbf{b}^{\mathsf{T}}\mathbf{z} \\ \text{s.t.} \quad & u(\mathrm{h}(\mathbf{x},\boldsymbol{\xi})) - r + \mathbf{z}^{\mathsf{T}}\psi(\boldsymbol{\xi}) \ \geq \ 0 \ , \ \forall \boldsymbol{\xi} \in \mathcal{S} \\ & \mathbf{z} \in \mathcal{K} \end{aligned}$$

-

Solving the DRSP

- Given that $\exists F \in \mathcal{D}(\psi)$ that strictly satisfies all moment constraints, then one can apply duality and solve the equivalent problem:

$$
\begin{aligned}
\underset{\mathbf{x},\ r,\mathbf{z}}{\text{maximize}} \quad & r - \mathbf{b}^\mathsf{T}\mathbf{z} \\
\text{s.t.} \quad & \alpha_k \mathsf{h}(\mathbf{x},\boldsymbol{\xi}) + \beta_k - r + \mathbf{z}^\mathsf{T}\psi(\boldsymbol{\xi}) \ \geq\ 0\ ,\ \forall\, \boldsymbol{\xi} \in \mathcal{S}\ ,\ \forall\, k \\
& \mathbf{z} \in \mathcal{K}\ , \quad \mathbf{x} \in \mathcal{X}
\end{aligned}
$$

Solving the DRSP

- Given that $\exists F \in \mathcal{D}(\psi)$ that strictly satisfies all moment constraints, then one can apply duality and solve the equivalent problem:

$$\begin{array}{ll} \underset{\mathbf{x},\; r,\mathbf{z}}{\text{maximize}} & r - \mathbf{b}^{\mathsf{T}}\mathbf{z} \\ \text{s.t.} & \alpha_k \mathrm{h}(\mathbf{x},\boldsymbol{\xi}) + \beta_k - r + \mathbf{z}^{\mathsf{T}}\psi(\boldsymbol{\xi}) \;\geq\; 0\;,\; \forall\, \boldsymbol{\xi} \in \mathcal{S}\;,\; \forall\, k \\ & \mathbf{z} \in \mathcal{K}\;, \quad \mathbf{x} \in \mathcal{X} \end{array}$$

- Applying the robust optimization analysis, we can reformulate each constraints through duality:

$$\min_{\boldsymbol{\xi} \in \mathcal{S}} \;\; \alpha_k \mathrm{h}(\mathbf{x},\boldsymbol{\xi}) + \beta_k - r + \mathbf{z}^{\mathsf{T}}\psi(\boldsymbol{\xi}) \;\geq\; 0$$

Outline

Distributionally Robust Portfolio Optimization

Consider the case of portfolio optimization (with call options):

$$\begin{array}{ll} \max_{\mathbf{x},\mathbf{y}}. & \inf_{F\in\mathcal{D}} \mathbb{E}_F[u(\mathrm{h}(\mathbf{x},\mathbf{y},\boldsymbol{\xi}))] \\ \text{s.t.} & \sum_i x_i + y_i \leq B \\ & \mathbf{x} \geq 0\,, \quad \mathbf{y} \geq 0 \end{array}$$

where x_i = how much is invested in stock i with future return ξ_i, y_i = amount invested in call option purchased on ξ_i, and

$$\begin{array}{ll} \mathrm{h}(\mathbf{x},\mathbf{y},\boldsymbol{\xi}) := \max_{\mathbf{z}}. & \boldsymbol{\xi}^T\mathbf{x} + \sum_i (\xi_i - \bar{\xi}_i) z_i \\ \text{s.t.} & 0 \leq \mathbf{z} \leq \mathbf{y} \end{array}$$

Distributionally Robust Portfolio Optimization

Consider the case of portfolio optimization (with call options):

$$\begin{aligned} \max_{\mathbf{x},\mathbf{y}}. \quad & \inf_{F\in\mathcal{D}} \mathbb{E}_F[u(\mathrm{h}(\mathbf{x},\mathbf{y},\boldsymbol{\xi}))] \\ \text{s.t.} \quad & \sum_i x_i + y_i \leq B \\ & \mathbf{x} \geq 0\,, \quad \mathbf{y} \geq 0 \end{aligned}$$

where x_i = how much is invested in stock i with future return ξ_i, y_i = amount invested in call option purchased on ξ_i, and

$$\begin{aligned} \mathrm{h}(\mathbf{x},\mathbf{y},\boldsymbol{\xi}) := \max_{\mathbf{z}}. \quad & \boldsymbol{\xi}^T\mathbf{x} + \sum_i (\xi_i - \bar{\xi}_i) z_i \\ \text{s.t.} \quad & 0 \leq \mathbf{z} \leq \mathbf{y} \end{aligned}$$

We bound the distributions with $\mathcal{D}_2(\gamma)$ and consider the support set described by $\mathcal{S} = \{\boldsymbol{\xi} | \mathbf{A}\boldsymbol{\xi} \leq \mathbf{b}\}$

DRSP reduces to SDP

In this context, the robust constraint

$$\min_{\boldsymbol{\xi}\in\mathcal{S}} \;\; \alpha_k \mathrm{h}(\mathbf{x},\boldsymbol{\xi}) + \beta_k + \mathbf{z}^{\mathsf{T}}\psi(\boldsymbol{\xi}) \;\geq\; 0$$

becomes:

$$\underset{\boldsymbol{\xi}\in\mathcal{S}}{\text{minimize}} \;\; \max_{0\leq\mathbf{z}\leq\mathbf{y}} \alpha_k(\boldsymbol{\xi}^T\mathbf{x}+\sum_i(\xi_i-\bar{\xi}_i)z_i)+\beta_k-r+2\mathbf{p}^T\boldsymbol{\xi}+\boldsymbol{\xi}^T\mathbf{Q}\boldsymbol{\xi}-2\hat{\boldsymbol{\mu}}^T\mathbf{Q}\boldsymbol{\xi}$$

DRSP reduces to SDP

In this context, the robust constraint

$$\min_{\boldsymbol{\xi}\in\mathcal{S}} \ \alpha_k \mathsf{h}(\mathbf{x},\boldsymbol{\xi}) + \beta_k + \mathbf{z}^{\mathsf{T}}\psi(\boldsymbol{\xi}) \ \geq \ 0$$

becomes:

$$\operatorname*{minimize}_{\boldsymbol{\xi}\in\mathcal{S}} \ \ \max_{0\leq\mathbf{z}\leq\mathbf{y}} \alpha_k(\boldsymbol{\xi}^T\mathbf{x}+\sum_i(\xi_i-\bar{\xi}_i)z_i)+\beta_k-r+2\mathbf{p}^T\boldsymbol{\xi}+\boldsymbol{\xi}^T\mathbf{Q}\boldsymbol{\xi}-2\hat{\boldsymbol{\mu}}^T\mathbf{Q}\boldsymbol{\xi}$$

By Sion's minimax and with use of duality we get the problem:

$$\max_{0\leq\mathbf{z}\leq\mathbf{y}}. \max_{\lambda\geq 0} \min_{\boldsymbol{\xi}} \ \ \alpha_k(\boldsymbol{\xi}^T\mathbf{x}+\sum_i(\xi_i-\bar{\xi}_i)z_i)+\beta_k \ \ \ldots \ -2\hat{\boldsymbol{\mu}}^T\mathbf{Q}\boldsymbol{\xi}+\lambda^T(\mathbf{A}\boldsymbol{\xi}-\mathbf{b})$$

DRSP reduces to SDP

In this context, the robust constraint

$$\min_{\boldsymbol{\xi}\in\mathcal{S}} \ \alpha_k \mathrm{h}(\mathbf{x},\boldsymbol{\xi}) + \beta_k + \mathbf{z}^{\mathsf{T}}\psi(\boldsymbol{\xi}) \ \geq \ 0$$

becomes:

$$\underset{\boldsymbol{\xi}\in\mathcal{S}}{\text{minimize}} \ \max_{0\leq\mathbf{z}\leq\mathbf{y}} \alpha_k(\boldsymbol{\xi}^T\mathbf{x}+\sum_i(\xi_i-\bar{\xi}_i)z_i)+\beta_k-r+2\mathbf{p}^T\boldsymbol{\xi}+\boldsymbol{\xi}^T\mathbf{Q}\boldsymbol{\xi}-2\hat{\boldsymbol{\mu}}^T\mathbf{Q}\boldsymbol{\xi}$$

By Sion's minimax and with use of duality we get the problem:

$$\max_{0\leq\mathbf{z}\leq\mathbf{y}}.\max_{\lambda\geq 0}\min_{\boldsymbol{\xi}} \ \alpha_k(\boldsymbol{\xi}^T\mathbf{x}+\sum_i(\xi_i-\bar{\xi}_i)z_i)+\beta_k \ \ \ldots \ -2\hat{\boldsymbol{\mu}}^T\mathbf{Q}\boldsymbol{\xi}+\lambda^T(\mathbf{A}\boldsymbol{\xi}-\mathbf{b})$$

which value is greater or equal to zero iff

$$\begin{bmatrix} \mathbf{Q} & (\alpha_k(\mathbf{x}+\mathbf{z})+2\mathbf{p}-2\mathbf{Q}\hat{\boldsymbol{\mu}}+\mathbf{A}^T\lambda)/2 \\ \ldots & -\alpha_k\bar{\boldsymbol{\xi}}^T\mathbf{z}+\beta_k-\lambda^T\mathbf{b}-r \end{bmatrix} \succeq 0$$

for some $0\leq\mathbf{z}\leq\mathbf{y}$ and $\lambda\geq 0$.

DRSP reduces to SDP II

When the dust settles down we get:

$$\max_{\mathbf{x},\mathbf{y}\; r,\mathbf{Q},\mathbf{P},\mathbf{p},s,\; \mathbf{z}^k,\lambda^k} \quad r - \left((1+\gamma_2)\hat{\Sigma} - \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^{\mathsf{T}}\right) \bullet \mathbf{Q} - \hat{\Sigma} \bullet \mathbf{P} - 2\hat{\boldsymbol{\mu}}^T \mathbf{p} - \gamma_1 s$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{P} & \mathbf{p} \\ \mathbf{p}^T & s \end{bmatrix} \succeq 0\,, \quad \mathbf{Q} \succeq 0$$

$$\begin{bmatrix} \mathbf{Q} & (\alpha_k(\mathbf{x}+\mathbf{z}^k) + 2\mathbf{p} - 2\mathbf{Q}\hat{\boldsymbol{\mu}} + \mathbf{A}^T\lambda)/2 \\ \dots & -\alpha_k\bar{\xi}^T\mathbf{z}^k + \beta_k - \mathbf{b}^T\lambda^k - r] \end{bmatrix} \succeq 0\,,\ \forall\, k$$

$$\sum_i x_i + y_i \le B$$

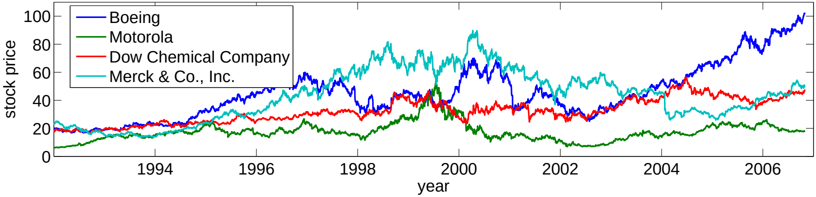
$$0 \le \mathbf{z}^k \le \mathbf{y}\,,\ \forall\, k$$

$$\lambda^k \ge 0\,, \forall\, k$$

$$\mathbf{x} \ge 0\,, \quad \mathbf{y} \ge 0$$

Experiments in Portfolio Optimization

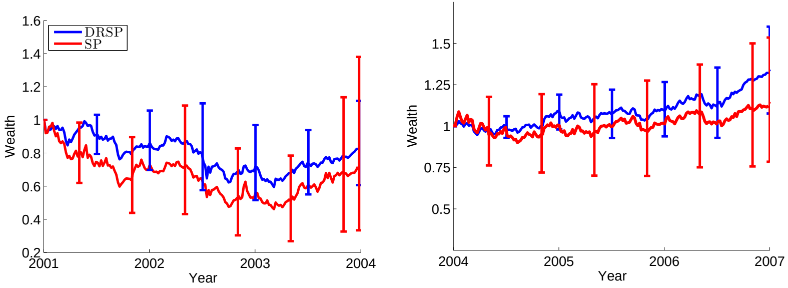
30 stocks tracked over years 1992-2007 using Yahoo! Finance



How does the robust solution perform compared to stochastic programming solution?

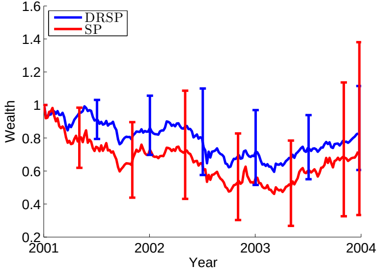
$$\mathcal{D} = \mathcal{D}_2(\gamma) \qquad \text{vs.} \qquad \mathcal{D} = \{\hat{F}\}$$

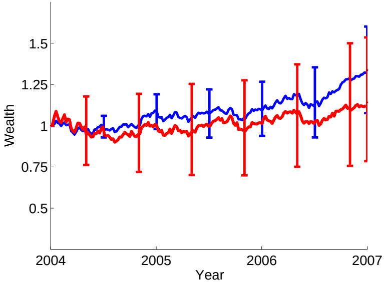
Wealth Evolution for 300 Experiments



- 10% and 90% percentiles are indicated periodically.

Wealth Evolution for 300 Experiments





- 10% and 90% percentiles are indicated periodically.
- 79% of time, the DRSP outperformed the stoch. prog. model
- 67% improvement on average using DRSP with $\mathcal{D}_2(\gamma)$

Outline

Are Airlines Adventurous in their Fleet Acquisition?

- Fleet composition is a difficult decision problem:
  - Fleet contracts are signed 10 to 20 years ahead of schedule.
  - Many factors are still unknown at that time:
    e.g., passenger demand, fuel prices, etc.

Are Airlines Adventurous in their Fleet Acquisition?

- Fleet composition is a difficult decision problem:
  - Fleet contracts are signed 10 to 20 years ahead of schedule.
  - Many factors are still unknown at that time:
    e.g., passenger demand, fuel prices, etc.
- Yet, most airline companies sign these contracts based on a single scenario of what the future may be.

Are Airlines Adventurous in their Fleet Acquisition?

- Fleet composition is a difficult decision problem:
  - Fleet contracts are signed 10 to 20 years ahead of schedule.
  - Many factors are still unknown at that time:
    e.g., passenger demand, fuel prices, etc.
- Yet, most airline companies sign these contracts based on a single scenario of what the future may be.
- Are airlines companies at risk of not being profitable in long run ?

Mathematical Formulation for Fleet Mix Optimization

The fleet composition problem is a stochastic mixed integer LP

$$\underset{\mathbf{x}}{\text{maximize}} \quad \mathbb{E}\,[-\underbrace{\mathbf{o}^\top\mathbf{x}}_{\text{ownership cost}} + \underbrace{h(\mathbf{x}, \tilde{\mathbf{p}}, \tilde{\mathbf{c}}, \tilde{\mathbf{L}})}_{\text{future profits}}\,] \;,$$

Fleet mix → $\mathbf{x}$

Mathematical Formulation for Fleet Mix Optimization

The fleet composition problem is a stochastic mixed integer LP

$$\underset{\mathbf{x}}{\text{maximize}} \quad \mathbb{E}[-\underbrace{\mathbf{o}^T\mathbf{x}}_{\text{ownership cost}} + \underbrace{h(\mathbf{x}, \tilde{\mathbf{p}}, \tilde{\mathbf{c}}, \tilde{\mathbf{L}})}_{\text{future profits}}] \;,$$

Fleet mix → $\mathbf{x}$

with $h(\mathbf{x}, \tilde{\mathbf{p}}, \tilde{\mathbf{c}}, \tilde{\mathbf{L}}) :=$

$$\max_{z\geq 0, y\geq 0, w} \quad \sum_k \Big(\sum_i \overbrace{\tilde{p}_i^k w_i^k}^{\text{flight profit}} - \overbrace{\tilde{c}_k (z_k - x_k)^+}^{\text{rental cost}} + \overbrace{\tilde{L}_k (x_k - z_k)^+}^{\text{lease revenue}} \Big)$$

$$\text{s.t.} \quad w_i^k \in \{0,1\}\,,\ \forall\, k, \forall\, i \quad \& \quad \sum_k w_i^k = 1\,,\ \forall\, i \quad \Big\}\ \text{Cover}$$

$$y^k_{g\in\text{in}(v)} + \sum_{i\in\text{arr}(v)} w_i^k = y^k_{g\in\text{out}(v)} + \sum_{i\in\text{dep}(v)} w_i^k\,,\ \forall\, k, \forall\, v \quad \Big\}\ \text{Balance}$$

$$z_k = \sum_{v\in\{v|\text{time}(v)=0\}} \big(y^k_{g\in\text{in}(v)} + \sum_{i\in\text{arr}(v)} w_i^k\big)\,, \quad \forall k \quad \Big\}\ \text{Count}$$

The Robust Mean Value Problem

Theorem

Given any set $\mathcal{D}(\psi)$, the solution of

$$(RMVP) \quad \underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \inf_{F\in\mathcal{D}(\psi)} h(\mathbf{x}, \mathbb{E}_F[\boldsymbol{\xi}])$$

is optimal with respect to

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \inf_{F\in\mathcal{D}(\psi)} \quad \mathbb{E}_F[h(\mathbf{x}, \boldsymbol{\xi})] \ .$$

The Robust Mean Value Problem

Theorem

Given any set $\mathcal{D}(\psi)$, the solution of

$$(RMVP) \quad \underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \inf_{F\in\mathcal{D}(\psi)} h(\mathbf{x}, \mathbb{E}_F[\boldsymbol{\xi}])$$

is optimal with respect to

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \inf_{F\in\mathcal{D}(\psi)} \mathbb{E}_F[h(\mathbf{x}, \boldsymbol{\xi})] \ .$$

Proof: Since $\mathrm{h}(\mathbf{x}, \cdot)$ is convex, any feasible distribution $F \in \mathcal{D}(\psi)$ is more optimistic than the more concentrated distribution $\delta_{\mathbb{E}_F[\boldsymbol{\xi}]} \in \mathcal{D}(\psi)$:

$$\mathbb{E}_F[\mathrm{h}(\mathbf{x}, \boldsymbol{\xi})] \geq \mathrm{h}(\mathbf{x}, \mathbb{E}_F[\boldsymbol{\xi}]) = \mathbb{E}_{\delta_\mu}[\mathrm{h}(\mathbf{x}, \boldsymbol{\xi})]$$

Insights

- Resolving the mean value is more important than resolving the distribution
- The robustness of the mean value problem holds for any two-stage stochastic linear programming problem with uncertainty in objective and a risk neutral attitude
- There is a need for a tool to estimate the value of stochastic modeling

Experiments in Fleet Mix Optimization

We experimented with three test cases :

1. 3 types of aircrafts, 84 flights
2. 4 types of aircrafts, 240 flights
3. 13 types of aircrafts, 535 flights

Experiments in Fleet Mix Optimization

We experimented with three test cases :

1. 3 types of aircrafts, 84 flights
2. 4 types of aircrafts, 240 flights
3. 13 types of aircrafts, 535 flights

Results (known μ & Σ, unknown F):

Test cases	Solution time		Regret for RMVP based on SP	Worst-case regret for RMVP solution
	RMVP	SP (100 scen.)		
#1	0.5 s	3 min	0.02%	$\leq$ 6%
#2	1 s	20 min	0.002%	$\leq$ 1%
#3	30 s	6 h	0.003%	$\leq$ 8%

Experiments in Fleet Mix Optimization

We experimented with three test cases :

1. 3 types of aircrafts, 84 flights
2. 4 types of aircrafts, 240 flights
3. 13 types of aircrafts, 535 flights

Results (known μ & Σ, unknown F):

Test cases	Solution time		Regret for RMVP based on SP	Worst-case regret for RMVP solution
	RMVP	SP (100 scen.)		
#1	0.5 s	3 min	0.02%	$\leq$ 6%
#2	1 s	20 min	0.002%	$\leq$ 1%
#3	30 s	6 h	0.003%	$\leq$ 8%

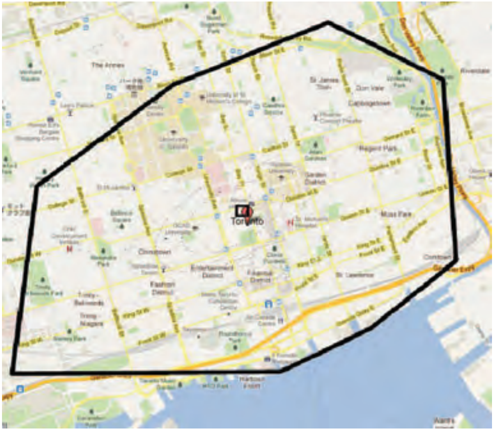
Finding:

- It would wasteful for these airline companies to invest more than 8% of profits in development of a stochastic model

Outline

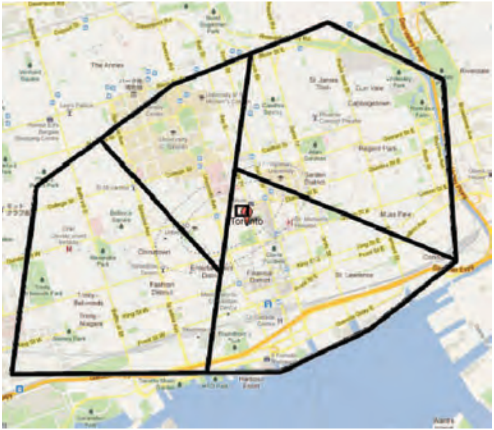
Parcel delivery in Toronto

- Divide downtown Toronto region into K subregions, each serviced by a different vehicle, so that the total workload be most evenly distributed among the fleet



Parcel delivery in Toronto

- Divide downtown Toronto region into K subregions, each serviced by a different vehicle, so that the total workload be most evenly distributed among the fleet



Parcel delivery in Toronto

- Divide downtown Toronto region into K subregions, each serviced by a different vehicle, so that the total workload be most evenly distributed among the fleet

Parcel delivery in Toronto

- Divide downtown Toronto region into K subregions, each serviced by a different vehicle, so that the total workload be most evenly distributed among the fleet

Partitioning under a Known Demand Distribution

- Given a distribution F of demand points, we would like to partition so that the workload of the busiest driver is as small as possible

$$\min_{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}} \left\{ \max_i \mathbb{E}[TSP(\{\xi_1,\xi_2,\ldots,\xi_N\}\cap\mathcal{R}_i)] \right\} \quad ,$$

Partitioning under a Known Demand Distribution

- Given a distribution F of demand points, we would like to partition so that the workload of the busiest driver is as small as possible

$$\min_{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}} \left\{ \max_i \mathbb{E}[TSP(\{\xi_1,\xi_2,\ldots,\xi_N\}\cap\mathcal{R}_i)] \right\} ,$$

- When the number of demand points in a region is large, Beardwood et al. (1959) tells us that

$$TSP(\{\xi_1,\xi_2,\ldots,\xi_N\}\cap\mathcal{R}_i) = \beta\sqrt{N}\iint_{\mathcal{R}_i}\sqrt{f(\xi)}d\xi + o(\sqrt{N})$$

Partitioning under a Known Demand Distribution

- Given a distribution F of demand points, we would like to partition so that the workload of the busiest driver is as small as possible

$$\min_{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}} \left\{ \max_i \mathbb{E}[TSP(\{\xi_1,\xi_2,\ldots,\xi_N\}\cap\mathcal{R}_i)] \right\} \quad ,$$

- When the number of demand points in a region is large, Beardwood et al. (1959) tells us that

$$TSP(\{\xi_1,\xi_2,\ldots,\xi_N\}\cap\mathcal{R}_i) = \beta\sqrt{N}\iint_{\mathcal{R}_i}\sqrt{f(\xi)}d\xi + o(\sqrt{N})$$

- Hence, Carlsson et al. (2007) suggest partitioning according to

$$\min_{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}} \left\{ \max_i \iint_{\mathcal{R}_i}\sqrt{f(\xi)}d\xi \right\} \quad ,$$

Partitioning under a Known Demand Distribution

- Given a distribution F of demand points, we would like to partition so that the workload of the busiest driver is as small as possible

$$\min_{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}} \left\{\max_i \mathbb{E}[TSP(\{\xi_1,\xi_2,\ldots,\xi_N\}\cap\mathcal{R}_i)]\right\} ,$$

- When the number of demand points in a region is large, Beardwood et al. (1959) tells us that

$$TSP(\{\xi_1,\xi_2,\ldots,\xi_N\}\cap\mathcal{R}_i) = \beta\sqrt{N}\iint_{\mathcal{R}_i}\sqrt{f(\xi)}d\xi + o(\sqrt{N})$$

- Hence, Carlsson et al. (2007) suggest partitioning according to

$$\min_{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}} \left\{\max_i \iint_{\mathcal{R}_i}\sqrt{f(\xi)}d\xi\right\} ,$$

- What can we do if we only have historical samples ?

Distributionally Robust Partitioning

- Given $\mathcal{D}_2(\gamma)$, we partition so that the largest workload over the worst distribution of demand points is as small as possible

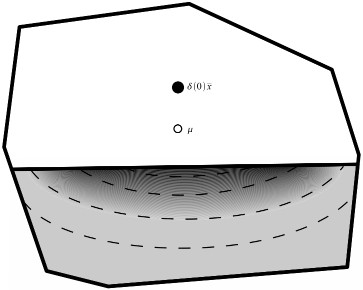
$$\min_{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}} \sup_{F\in\mathcal{D}_2(\gamma)} \left\{ \max_i \iint_{\mathcal{R}_i} \sqrt{f(\xi)}d\xi \right\} \quad ,$$

Distributionally Robust Partitioning

- Given $\mathcal{D}_2(\gamma)$, we partition so that the largest workload over the worst distribution of demand points is as small as possible

$$\underset{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}}{\text{min.}} \quad \sup_{F\in\mathcal{D}_2(\gamma)} \left\{ \max_i \iint_{\mathcal{R}_i} \sqrt{f(\xi)}d\xi \right\} \quad ,$$

- A side product will be to characterize for any partition what is a worst-case distribution of demand locations



Evaluating the Worst-case Performance

- For a given partition, one needs to evaluate the worst-case load for each region i:

$$\max_{F \in \mathcal{D}_2(\gamma)} \iint_{\mathcal{R}_i} \sqrt{f(\xi)} d\xi$$

Evaluating the Worst-case Performance

- For a given partition, one needs to evaluate the worst-case load for each region i:

$$\max_{F \in \mathcal{D}_2(\gamma)} \iint_{\mathcal{R}_i} \sqrt{f(\xi)} d\xi$$

- Using duality theory, we can show that it is equivalent to:

$$\begin{aligned}
\min_{r,\mathbf{Q},\mathbf{P},\mathbf{p},t,\lambda}. \quad & 1/4 \iint_{\mathcal{R}_i} \frac{1}{r + \boldsymbol{\xi}^T(2\mathbf{p} - 2\mathbf{Q}\hat{\boldsymbol{\mu}}) + \boldsymbol{\xi}^T\mathbf{Q}\boldsymbol{\xi}} d\boldsymbol{\xi} \\
& + r + ((1+\gamma_2)\hat{\Sigma} - \hat{\mu}\hat{\mu}^T) \bullet \mathbf{Q} + \hat{\Sigma} \bullet \mathbf{P} + 2\hat{\boldsymbol{\mu}}^T\mathbf{p} + \gamma_1 s \\
\text{s.t.} \quad & \begin{bmatrix} \mathbf{P} & \mathbf{p} \\ \mathbf{p}^T & s \end{bmatrix} \succeq 0 \\
& \begin{bmatrix} \mathbf{Q} & (2\mathbf{p} - 2\mathbf{Q}\hat{\boldsymbol{\mu}} + \mathbf{A}^T\lambda)/2 \\ \cdots & r - \mathbf{b}^T\lambda \end{bmatrix} \succeq 0 \qquad \lambda \geq 0 \ ,
\end{aligned}$$

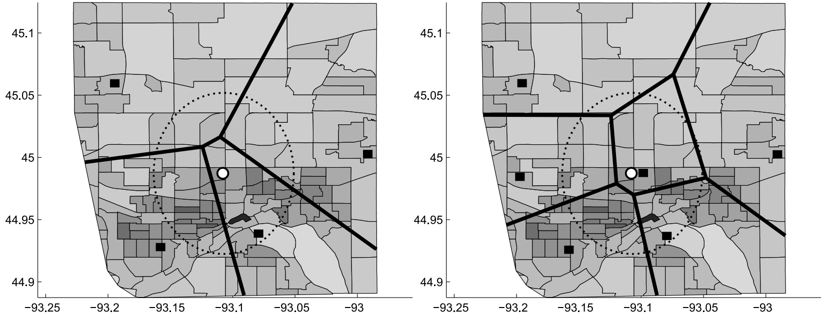
where $\mathbf{A}\boldsymbol{\xi} \leq \mathbf{b}$ describes the polygon $\mathcal{T}$.

- This problem can be formulated as an SDP after approximation of the integral by a finite sum.

Ramsey County Partition for FedEx Deliveries

Assigning each described region to a FedEx vehicle minimizes the worst-case tour of any one of them in a single day.

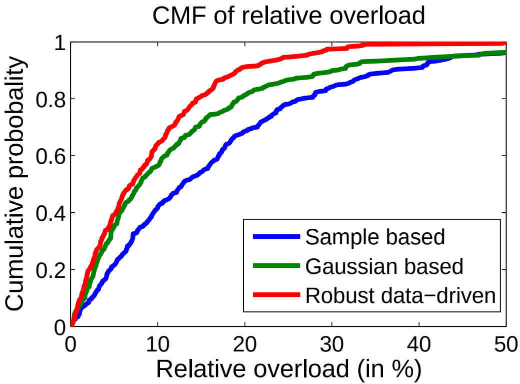
Optimal power diagrams based partitions



n=4 n=6

Distributionally Robust 2-Partitioning

We simulated three partition schemes on a set of randomly generated parcel delivery problems where the territory needed to be divided into two regions and the demand is drawn from a mixture of truncated Gaussian distribution



Outline

Conclusion

- DRSP with $\mathcal{D}(\psi)$, which imposes concentration properties, can be reformulated as a standard finite dimensional robust optimization problem
- There are cases where the robust mean value problem generates distributionally robust decisions
- DRO approach can easily lead to better decisions (both in terms of lower risk and higher returns) than an approach based on inaccurate stochastic model
- Some tools can help estimate how much identifying the true distribution is worth

Future Work

- Can a DRSP that also accounts for higher order moments be made tractable (polynomial optimization)? Would it help?

Future Work

- Can a DRSP that also accounts for higher order moments be made tractable (polynomial optimization)? Would it help?
- How to deal with ambiguity about one's risk attitude ?

$$\begin{array}{ll} \underset{t, x \in \mathcal{X}}{\text{maximize}} & t \\ \text{s.t.} & \mathbb{E}[u(\mathrm{h}(\mathbf{x}, \boldsymbol{\xi}))] \geq u(t), \ \forall\, u \in \mathcal{U} \end{array}$$

where $\mathcal{U}$ is the set of all risk averse utility functions that are coherent with a finite set of known preferences

Future Work

- Can a DRSP that also accounts for higher order moments be made tractable (polynomial optimization)? Would it help?
- How to deal with ambiguity about one's risk attitude ?

$$\begin{array}{ll} \underset{t, x \in \mathcal{X}}{\text{maximize}} & t \\ \text{s.t.} & \mathbb{E}[u(\mathsf{h}(\mathbf{x}, \boldsymbol{\xi}))] \geq u(t), \ \forall\, u \in \mathcal{U} \end{array}$$

  where $\mathcal{U}$ is the set of all risk averse utility functions that are coherent with a finite set of known preferences
- Can lesson's learned from DRO be usefull for multi-objective problems ?

Bibliography

Armbruster, B., E. Delage. 2011. Decision making under uncertainty when preference information is incomplete. Working paper.

Bertsimas, D., X. V. Doan, K. Natarajan, C. P. Teo. 2010. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research* **35**(3) 580–602.

Carlsson, J. G., E. Delage. 2011. Robust partitioning for stochastic multi-vehicle routing. Working paper.

Delage, E., S. Arroyo, Y. Ye. 2011. The value of stochastic modeling in two-stage stochastic programs with cost uncertainty. Working paper.

Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* **58**(3) 595–612.

Natarajan, K., M. Sim, J. Uichanco. 2010. Tractable Robust Expected Utility and Risk Models for Portfolio Optimization. *Mathematical Finance* **20**(4) 695–731.

Questions & Comments ...

... Thank you!