The Value of Distribution Information in Distributionally Robust Optimization

Erick Delage
CRC in Decision Making under Uncertainty
HEC Montréal

(joint work with S. Arroyo (The Boeing Cie.) and Y. Ye (Stanford University))

Monday, November 9th, 2015
Let’s consider a decision model that accounts for uncertainty:

\[ \text{(SP)} \quad \max_{x \in \mathcal{X}} \mathbb{E} [h(x, \xi)] \]

- \( x \) is a vector of decision variables in \( \mathbb{R}^n \)
- \( \xi \) is a vector of uncertain parameters in \( \mathbb{R}^m \)
- \( h(x, \xi) \) is a profit function

To find an optimal solution, one must develop a stochastic model and solve the associated stochastic program.
Difficulties in choosing a distribution model

- Developing an accurate stochastic model requires heavy engineering efforts and might even be impossible.

- This motivates the use of a distributionally robust optimization model:

  \[(\text{DRO}) \quad \max_{x \in X} \inf_{F \in \mathcal{D}} \mathbb{E}_F[h(x, \xi)] .\]

  where \( \mathcal{D} \) captures exactly what is known of the distribution.
Many methods have been proposed to convert i.i.d. samples \( \{\xi_i\}_{i=1}^M \) into confidence regions for distributions:

- **Hypothesis testing methods:** [(Bertsimas et al., 2015)]

\[
S \& \{\xi_i\}_{i=1}^M \rightarrow D := \{F \mid \exists \theta, \psi(F) = \theta, T_\theta(\{\xi_i\}) \leq \gamma(M)\}
\]

- **Moment based method:**
  [(Delage and Ye, 2010), (Wiesemann et al., 2014)]

\[
S \& \{\xi_i\}_{i=1}^M \rightarrow D_{\text{moment}} := \left\{ F \left| \begin{array}{l}
\mathbb{P}(\xi \in S) = 1 \\
\|\mathbb{E}[\xi] - \hat{\mu}\|_F^2 \leq O\left(\frac{\log(1/\delta)}{M}\right) \\
\mathbb{E}[(\xi - \hat{\mu})(\xi - \hat{\mu})^T] \preceq \left(1 + O\left(\sqrt{\frac{\log(1/\delta)}{M}}\right)\right) \hat{\Sigma}
\end{array}\right. \right\}
\]

- **Distance/divergence based methods:**
  [(Ben-Tal et al., 2013), (Mohajerin Esfahani et al., 2015)]

\[
S \& \{\xi_i\}_{i=1}^M \rightarrow D := \{F \mid d(F, \hat{F}) \leq \gamma(M)\}
\]
Consider that there are two urns in front of you. The two urns contain 100 BLUE and RED balls in unknown proportions.

Choose among the following three gambles:

- **Gamble A**: If you draw a BLUE ball from urn #1, then you win 180$, otherwise you win 20$
- **Gamble B**: If you draw a BLUE ball from urn #1, then you win 200$, otherwise you win nothing
- **Gamble C**: If you draw a BLUE ball from urn #2, then you win 100$, otherwise you win nothing

<table>
<thead>
<tr>
<th>Gamble A</th>
<th>Urn #2</th>
<th>Gamble B</th>
<th>Urn #2</th>
<th>Gamble C</th>
<th>Urn #2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blue</td>
<td>Urn #1</td>
<td>Blue</td>
<td>Urn #1</td>
<td>Blue</td>
</tr>
<tr>
<td>Urn #1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blue</td>
<td>180</td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Red</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urn #2</td>
<td>Blue</td>
<td>Red</td>
<td>Blue</td>
<td>Red</td>
<td></td>
</tr>
<tr>
<td>Blue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>180</td>
<td>20</td>
<td>200</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Physical ambiguity in Two Urns experiment

Consider that there are two urns in front of you. The two urns contain 100 BLUE and RED balls in unknown proportions.

Choose among the following three gambles:

- **Gamble A**: If you draw a BLUE ball from urn #1, then you win 180$, otherwise you win 20$
- **Gamble B**: If you draw a BLUE ball from urn #1, then you win 200$, otherwise you win nothing
- **Gamble C**: If you draw a BLUE ball from urn #2, then you win 100$, otherwise you win nothing

Distributionally robust optimization model is:

$$\max_{x \in \{0,1\}^3, x_A+x_B+x_C=1} \quad \min_{p \in [0,1]^2} (20 + 160p_1)x_A + (200p_1)x_B + (100p_2)x_C$$
How can one quantify the value of distribution information?

- In Two Urns experiment, what is the value of knowing the proportion of balls in either urn #1 or #2?
- In data-driven problems, what is the value of acquiring/processing more data?

This might serve many purposes:

- Indicate whether it is worth investing in acquisition of additional data
- Guide the type of data that should be acquired
Outline

1. Introduction
2. Three Different Measures
3. Some Theoretical Properties
4. Fleet Mix Optimization
5. Conclusion & Future Work
Outline

1. Introduction
2. Three Different Measures
3. Some Theoretical Properties
4. Fleet Mix Optimization
5. Conclusion & Future Work
Three possible measures

Let $\mathcal{O}$ be set of possible information that can be made and $\mathcal{D}(o)$ describe the update rule for the distribution set, such that $\mathcal{D} = \bigcup_{o \in \mathcal{O}} \mathcal{D}(o)$.

- **Worst-case value of information:**

$$WC-VDI(\mathcal{O}) = \min_{o \in \mathcal{O}} \max_{x \in \mathcal{X}} \min_{F \in \mathcal{D}(o)} \mathbb{E}_F[h(x, \xi)] - \max_{x \in \mathcal{X}} \min_{F \in \mathcal{D}} \mathbb{E}_F[h(x, \xi)]$$

- **Best-case value of information**

$$BC-VDI(\mathcal{O}) = \max_{o \in \mathcal{O}} \max_{x \in \mathcal{X}} \min_{F \in \mathcal{D}(o)} \mathbb{E}_F[h(x, \xi)] - \max_{x \in \mathcal{X}} \min_{F \in \mathcal{D}} \mathbb{E}_F[h(x, \xi)]$$

- **Worst-case regret of not using the information**

$$WCR-VDI(\mathcal{O}) = \max_{o \in \mathcal{O}} \left( \max_{x \in \mathcal{X}} \min_{F \in \mathcal{D}(o)} \mathbb{E}_F[h(x, \xi)] - \min_{F \in \mathcal{D}(o)} \mathbb{E}_F[h(x_0, \xi)] \right)$$

where $x_0 \in \text{arg max}_{x \in \mathcal{X}} \min_{F \in \mathcal{D}} \mathbb{E}_F[h(x, \xi)]$.
### Introduction

**Three Different Measures**

Some Theoretical Properties

Fleet Mix Optimization

Conclusion & Future Work

---

## Value of distribution information in Two Urns exp.

<table>
<thead>
<tr>
<th>Gamble A</th>
<th>Urn #2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blue</td>
<td>180</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Red</td>
<td>180</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gamble B</th>
<th>Urn #2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blue</td>
<td>200</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Red</td>
<td>200</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gamble C</th>
<th>Urn #2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blue</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Red</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

If we could count the balls of one urn, which one should it be?

- Based on WC-VDI:

\[
\text{WC-VDI(Urn#1)} = \min_{p_1 \in [0,1]} \max_{x \in \mathcal{X}} \min_{p_2 \in [0,1]} (20 + 160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\
- \max_{x \in \mathcal{X}} \min_{p \in [0,1]} (20 + 160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\
= \min_{p_1 \in [0,1]} \max_{x \in \mathcal{X}} (20 + 160p_1)x_A + (200p_1)x_B - 20 \\
= \max_{x \in \mathcal{X}} 20x_A - 20 = 0
\]
### Value of distribution information in Two Urns exp.

<table>
<thead>
<tr>
<th>Gamble</th>
<th>Urn #1</th>
<th>Urn #2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blue</td>
<td>Red</td>
</tr>
<tr>
<td>Gamble A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urn #1</td>
<td>Blue</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>Red</td>
<td>20</td>
</tr>
<tr>
<td>Gamble B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urn #1</td>
<td>Blue</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Red</td>
<td>0</td>
</tr>
<tr>
<td>Gamble C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urn #1</td>
<td>Blue</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Red</td>
<td>100</td>
</tr>
</tbody>
</table>

If we could count the balls of one urn, which one should it be?

- Based on WC-VDI:

\[
\begin{align*}
\text{WC-VDI(Urn#1)} &= 0 \\
\text{WC-VDI(Urn#2)} &= \min_{p_2 \in [0,1]} \max_{x \in \mathcal{X}} \min_{p_1 \in [0,1]} (20 + 160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\
& \quad - \max_{x \in \mathcal{X}} \min_{p \in [0,1]^2} (20 + 160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\
& = \min_{p_2 \in [0,1]} \max_{x \in \mathcal{X}} 20x_A + (100p_2)x_C - 20 = \max_{x \in \mathcal{X}} 20x_A - 20 = 0
\end{align*}
\]
If we could count the balls of one urn, which one should it be?

- Based on WC-VDI:

\[
\begin{align*}
\text{WC-VDI(Urn#1)} & = 20 - 20 = 0 \quad \text{(i.e. confirm no blue in urn #1.)} \\
\text{WC-VDI(Urn#2)} & = 20 - 20 = 0 \quad \text{(i.e. confirm no blue in urn #2.)}
\end{align*}
\]

- **Conclusion: Distribution information has no value!**
If we could count the balls of one urn, which one should it be?

- Based on BC-VDI:

\[
\text{BC-VDI(Urn\#1)} = \max_{p_1 \in [0, 1]} \max_{x \in X} \min_{p_2 \in [0, 1]} (20 + 160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\
- \max_{x \in X} \min_{p \in [0, 1]}^2 (20 + 160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\
= \max_{p_1 \in [0, 1]} \max_{x \in X} (20 + 160p_1)x_A + (200p_1)x_B - 20 \\
= \max_{x \in X} 180x_A + 200x_B - 20 = 180
\]
If we could count the balls of one urn, which one should it be?

- Based on BC-VDI:

\[
\text{BC-VDI}(\text{Urn#1}) = 180
\]
\[
\text{BC-VDI}(\text{Urn#2}) = \max_{p_2 \in [0,1]} \max_{x \in \mathcal{X}} \min_{p_1 \in [0,1]} (20 + 160p_1)x_A + (200p_1)x_B + (100p_2)x_C
\]
\[
- \min_{x \in \mathcal{X}} \max_{p \in [0,1]} (20 + 160p_1)x_A + (200p_1)x_B + (100p_2)x_C
\]
\[
= \max_{p_2 \in [0,1]} \max_{x \in \mathcal{X}} 20x_A + (100p_2)x_C - 20
\]
\[
= \max_{x \in \mathcal{X}} 20x_A + 100x_B - 20 = 80
\]
If we could count the balls of one urn, which one should it be?

- **Based on BC-VDI:**

  \[
  \text{BC-VDI}(\text{Urn#1}) = 200 - 20 = 180 \quad (\text{i.e. confirm all blue in urn #1.})
  \]

  \[
  \text{BC-VDI}(\text{Urn#2}) = 100 - 20 = 80 \quad (\text{i.e. confirm all blue in urn #2.})
  \]

- **Conclusion:** One should count the balls of Urn #1!
If we could count the balls of one urn, which one should it be?

- Based on BC-VDI:

  \[ BC-VDI(\text{Urn#1}) = 180 - 20 = 160 \] (i.e. confirm all blue in urn #1.)
  \[ BC-VDI(\text{Urn#2}) = 100 - 20 = 80 \] (i.e. confirm all blue in urn #2.)

- Conclusion: One should count the balls of Urn #1!
- Even when Gamble B is removed as an alternative!
If we could count the balls of one urn, which one should it be?

- Based on WCR-VDI:

\[
WCR-VDI(Urn\#1) = \max_{p_1 \in [0,1]} \max_{x \in \mathcal{X}} \min_{p_2 \in [0,1]} (20 + 160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\
- \min_{p_2 \in [0,1]} (20 + 160p_1) \cdot 1 + (200p_1) \cdot 0 + (100p_2) \cdot 0 \\
= \max_{p_1 \in [0,1]} \max_{x \in \mathcal{X}} (20 + 160p_1)x_A + (200p_1)x_B - (20 + 160p_1) \\
= \max_{p_1 \in [0,1]} \max_{x \in \mathcal{X}} (40p_1 - 20)x_B = \max_{x \in \mathcal{X}} 20x_B = 20
\]
Value of distribution information in Two Urns exp.

If we could count the balls of one urn, which one should it be?

- Based on WCR-VDI:

\[
\begin{align*}
\text{WCR-VDI(Urn#1)} &= 20 \\
\text{WCR-VDI(Urn#2)} &= \max_{p_2 \in [0,1]} \min_{p_1 \in [0,1]} \left( 20 + 160p_1 \right)x_A + (200p_1)x_B + (100p_2)x_C \\
&\quad - \min_{p_1 \in [0,1]} \left( 20 + 160p_1 \right) \cdot 1 + (200p_1) \cdot 0 + (100p_2) \cdot 0 \\
&= \max_{p_2 \in [0,1]} \max_{x \in \mathcal{X}} (20)x_A + (100p_2)x_B - 20 = \max_{x \in \mathcal{X}} 80x_B = 80
\end{align*}
\]

Delage et al.  Value of Distribution Information
Value of distribution information in Two Urns exp.

<table>
<thead>
<tr>
<th>Gamble A</th>
<th>Urn #2</th>
<th>Urn #1</th>
<th>Urn #2</th>
<th>Urn #1</th>
<th>Urn #2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blue</td>
<td>Red</td>
<td>Blue</td>
<td>Red</td>
<td>Blue</td>
</tr>
<tr>
<td>Urn #1</td>
<td>180</td>
<td>20</td>
<td>180</td>
<td>20</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>Blue</td>
<td>Red</td>
<td>Blue</td>
<td>Red</td>
<td>Blue</td>
</tr>
<tr>
<td>Gamble B</td>
<td>200</td>
<td>0</td>
<td>200</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Gamble C</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

If we could count the balls of one urn, which one should it be?

- Based on WCR-VDI:

  \[
  \text{WCR-VDI(Urn#1)} = 200 - 180 = 20 \quad (\text{i.e. confirm all blue in urn #1.})
  \]
  \[
  \text{WCR-VDI(Urn#2)} = 100 - 20 = 80 \quad (\text{i.e. confirm all blue in urn #2.})
  \]

- Conclusion: One should count the balls of Urn #2!
Introduction

Three Different Measures

Some Theoretical Properties

Fleet Mix Optimization

Conclusion & Future Work
An ordering of VDI measures

In Two Urns experiment, we noticed that

\[
\begin{align*}
\text{WC-VDI(Urn#1)} & \leq \text{WCR-VDI(Urn#1)} & \leq \text{BC-VDI(Urn#1)} \\
0 & \leq 20 & \leq 180 \\
\text{WC-VDI(Urn#2)} & \leq \text{WCR-VDI(Urn#2)} & \leq \text{BC-VDI(Urn#2)} \\
0 & \leq 80 & \leq 80
\end{align*}
\]

Lemma

It is generally the case that

\[
\text{WC-VDI}(\mathcal{O}) \leq \text{WCR-VDI}(\mathcal{O}) \leq \text{BC-VDI}(\mathcal{O}) .
\]
An ordering of VDI measures

Lemma

It is generally the case that

\[ WC-VDI(\mathcal{O}) \leq WCR-VDI(\mathcal{O}) \leq BC-VDI(\mathcal{O}). \]

Proof:

\[
WC-VDI(\mathcal{O}) = \min_{o_1 \in \mathcal{O}} \max_{x_1} \min_{F \in \mathcal{D}(o_1)} \mathbb{E}_{F}[h(x_1, \xi)] - \max_{x_2 \in \mathcal{X}} \min_{F \in \mathcal{D}} \mathbb{E}_{F}[h(x_2, \xi)]
\]

\[
= \min_{o_1 \in \mathcal{O}} \max_{x_1 \in \mathcal{X}} \min_{F \in \mathcal{D}(o_1)} \mathbb{E}_{F}[h(x_1, \xi)] - \min_{o_2 \in \mathcal{O}} \min_{F \in \mathcal{D}(o_2)} \mathbb{E}_{F}[h(x_0, \xi)]
\]

\[
= \max_{o_2 \in \mathcal{O}} \min_{o_1 \in \mathcal{O}} \max_{x_1 \in \mathcal{X}} \min_{F \in \mathcal{D}(o_1)} \mathbb{E}_{F}[h(x_1, \xi)] - \min_{F \in \mathcal{D}(o_2)} \mathbb{E}_{F}[h(x_0, \xi)]
\]

\[
\leq \max_{o_1 = o_2 \in \mathcal{O}} \min_{x_1 \in \mathcal{X}} \min_{F \in \mathcal{D}(o_1)} \mathbb{E}_{F}[h(x_1, \xi)] - \min_{F \in \mathcal{D}(o_2)} \mathbb{E}_{F}[h(x_0, \xi)]
\]

\[
= WCR-VDI(\mathcal{O})
\]
An ordering of VDI measures

Lemma

It is generally the case that

\[ \text{WC-VDI}(\mathcal{O}) \leq \text{WCR-VDI}(\mathcal{O}) \leq \text{BC-VDI}(\mathcal{O}). \]

Proof:

\[
\text{WCR-VDI}(\mathcal{O}) = \max_{o \in \mathcal{O}} \left\{ \max_{x_1 \in \mathcal{X}} \min_{F \in \mathcal{D}(o)} \mathbb{E}_F[h(x_1, \xi)] - \min_{F \in \mathcal{D}(o)} \mathbb{E}_F[h(x_0, \xi)] \right\}
\leq \max_{o_1 \in \mathcal{O}} \max_{o_2 \in \mathcal{O}} \left\{ \max_{x_1 \in \mathcal{X}} \min_{F \in \mathcal{D}(o_1)} \mathbb{E}_F[h(x_1, \xi)] - \min_{F \in \mathcal{D}(o_2)} \mathbb{E}_F[h(x_0, \xi)] \right\}
= \text{BC-VDI}(\mathcal{O})
\]
Many situations where VDI = 0 in the worst case

**Lemma**

If the feasible set \( \mathcal{X} \) is convex and compact and the profit function \( h(x, \xi) \) is concave in \( x \), then \( WC-VDI(\mathcal{O}) = 0 \).

Proof: Based on Sion’s minimax theorem we have that

\[
WC-VDI(\mathcal{O}) = \min_{o \in \mathcal{O}} \max_{x_1 \in \mathcal{X}} \min_{F \in \mathcal{D}(o)} \mathbb{E}_F[h(x_1, \xi)] - \max_{x_2 \in \mathcal{X}} \min_{F \in \mathcal{D}} \mathbb{E}_F[h(x_2, \xi)] \\
\leq \min_{o \in \mathcal{O}} \min_{F \in \mathcal{D}(o)} \max_{x_1 \in \mathcal{X}} \mathbb{E}_F[h(x_1, \xi)] - \max_{x_2 \in \mathcal{X}} \min_{F \in \mathcal{D}} \mathbb{E}_F[h(x_2, \xi)] \\
= \min_{F \in \mathcal{D}} \max_{x_1 \in \mathcal{X}} \mathbb{E}_F[h(x_1, \xi)] - \max_{x_2 \in \mathcal{X}} \min_{F \in \mathcal{D}} \mathbb{E}_F[h(x_2, \xi)] = 0.
\]
Some situations where VDI = 0 in the best case

**Theorem (Delage et al., 2014)**

Let the profit function $h(x, \xi)$ be convex in $\xi$, and let the distribution set be $D := \{F \mid \mathbb{E}_F[\xi] = \mu\}$. Then for any information sets of type $O := \{\gamma \in \mathbb{R}^+ \mid \mathbb{E}_F[\psi(\xi)] \leq \gamma\}$ where $\psi(\cdot)$ is a convex function and $D(o) \neq \emptyset$ for all $o \in O$, the value is zero even in the best case.

Proof:

$$BC\text{-VDI} = \max_{o \in O} \max_{x \in \mathcal{X}} \min_{F \in D(o)} \mathbb{E}_F[h(x, \xi)] - \max_{x \in \mathcal{X}} \min_{F \in D} \mathbb{E}_F[h(x, \xi)]$$

$$= \max_{o \in O} \max_{x \in \mathcal{X}} \mathbb{E}_F[h(x, \mu)] - \max_{x \in \mathcal{X}} \mathbb{E}_F[h(x, \mu)] = 0,$$

since Jensen’s inequality ensures that $\delta_\mu$ (i.e. the Dirac measure centred at $\mu$) always achieves a lower profit than any $F \in D$, and since this Dirac measure remains feasible when imposing that $\mathbb{E}_F[\psi(\xi)] \leq \gamma$. 

26 Delage et al. Value of Distribution Information
Evaluating worst-case regret is NP-hard

**Theorem (Delage et al., 2014)**

Evaluating BC-VDI($\mathcal{O}$) or WCR-VDI($\mathcal{O}$) exactly is NP-hard even when $h(x, \xi)$ is concave in $x$ and convex in $\xi$, and $\mathcal{D}_{\text{moment}}$ is used.

Sketch of proof:

- When the distribution information is perfect,

  $\text{WCR-VDI}(\mathcal{O}) = \max_{F \in \mathcal{D}} \max_{x \in \mathcal{X}} \mathbb{E}_F[h(x, \xi)] - \mathbb{E}_F[h(x_0, \xi)]$

  $= \max_{x \in \mathcal{X}} \max_{F \in \mathcal{D}} \mathbb{E}_F[h(x, \xi)] - \mathbb{E}_F[h(x_0, \xi)]$.

- Evaluating $\max_{F \in \mathcal{D}_{\text{moment}}} \mathbb{E}_F[h(x, \xi)]$ is NP-hard for

  $h(x, \xi) := \max_{y \in \mathbb{R}^m} c^T x + \xi^T y$

  s.t. $|y_i| \leq x$, $\forall y \in \{1, 2, \ldots, m\}$

  $a^T y = 0$. 

Delage et al. Value of Distribution Information
Theorem (Delage et al., 2014)

If the following conditions apply:

1. $D_{\text{moment}}$ is used with $S \subseteq \{\xi \mid \|\xi\|_1 \leq \rho\}$ and  
$$\|\mathbb{E}_F[\xi] - \hat{\mu}\|_{\hat{\Sigma}^{-1/2}}^2 \leq \gamma_1$$

2. $h(x, \xi)$ captures a two-stage linear program with cost uncertainty, i.e., $h(x, \xi) := \max_{y \in \mathcal{Y}(x)} c^T x + \xi^T C y$.

then an upper bound for $WCR-VDI(O)$ can be evaluated

$$WCR-VDI(O) \leq \min_{s \in \mathbb{R}, q \in \mathbb{R}^m} s + \hat{\mu}^T q + \sqrt{\gamma_1} \|\hat{\Sigma}^{1/2} q\|$$

subject to

$$s \geq \alpha(\rho e_i) - \rho e_i^T q , \forall i \in \{1, \ldots, m\}$$

$$s \geq \alpha(-\rho e_i) + \rho e_i^T q , \forall i \in \{1, \ldots, m\} ,$$

where $\alpha(\xi) = \max_{x \in X} h(x, \xi) - (c^T \bar{y}_0 + \xi^T C \bar{y}_0)$ for any $\bar{y}_0 \in \mathcal{Y}(x_0)$.
Fleet mix optimization is a difficult decision problem:
- Fleet contracts are signed 10 to 20 years ahead of schedule.
- Many factors are still unknown at that time: passenger demand, fuel prices, etc.

Yet, many airline companies sign these contracts based on a single scenario of what the future may be.

We first show that using the mean value of future profits as a scenario leads to the same solution as DRO with $\mathcal{D}_{\text{moment}}$ with known first moment

Can we do better by acquiring more information about the distribution?
The fleet composition problem is a stochastic mixed integer LP

\[
\text{maximize} \quad \mathbb{E} \left[ - o^T x + h(x, \tilde{p}, \tilde{c}, \tilde{L}) \right],
\]

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

\[
\max_{z \geq 0, y \geq 0, w} \sum_k \left( \sum_i \left( \tilde{p}_i^k w_i - \tilde{c}_k (z_k - x_k)^+ + \tilde{L}_k (x_k - z_k)^+ \right) \right)
\]

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

\[
y_{g \in \text{in}(v)} + \sum_{i \in \text{arr}(v)} w_i^k = y_{g \in \text{out}(v)} + \sum_{i \in \text{dep}(v)} w_i^k, \forall k, \forall v \} \text{ Balance}
\]

\[
z_k = \sum_{v \in \{ v | \text{time}(v) = 0 \}} (y_{g \in \text{in}(v)} + \sum_{i \in \text{arr}(v)} w_i^k), \forall k \} \text{ Count}
Experiments in fleet mix optimization

We experimented with three test cases:

1. 3 types of aircrafts, 84 flights, $\sigma_{\tilde{p}_i}/\mu_{\tilde{p}_i} \in [4\%, 53\%]$
2. 4 types of aircrafts, 240 flights, $\sigma_{\tilde{p}_i}/\mu_{\tilde{p}_i} \in [2\%, 20\%]$
3. 13 types of aircrafts, 535 flights, $\sigma_{\tilde{p}_i}/\mu_{\tilde{p}_i} \in [2\%, 58\%]$

Results:

<table>
<thead>
<tr>
<th>Test cases</th>
<th>WCR-VDI($\phi$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>$\leq 6%$</td>
</tr>
<tr>
<td>#2</td>
<td>$\leq 1%$</td>
</tr>
<tr>
<td>#3</td>
<td>$\leq 7%$</td>
</tr>
</tbody>
</table>

Conclusions:

- It’s wasteful in these problems to invest more than 7% of profits in acquisition of distribution information.
Conclusion & future work

- Need for tools that can estimate the value of distribution information
  - The most natural tools are computational intractable
  - Tractable upper bounds for value of perfect distribution information might be available and informative (e.g. fleet-mix optimization)

- Future work:
  - Develop tighter bounds for WCR-VDI($\mathcal{O}$) with perfect distribution information under $\mathcal{D}_{\text{moment}}$
  - Derive bounds for other distribution sets
  - Design simple procedures for characterizing $\mathcal{O}$ and $\mathcal{D}(o)$ and bounding WCR-VDI($\mathcal{O}$) in data-driven problem where information consists of samples


Questions & Comments ...