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

Decision making under uncertainty

Let's consider a decision model that accounts for uncertainty:

$$(\text{SP}) \qquad \underset{x \in \mathcal{X}}{\text{maximize}} \quad \mathbb{E}\left[h(x, \xi)\right]$$

- x is a vector of decision variables in $\mathbb{R}^n$
- ξ 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

$$(\text{DRO}) \qquad \underset{x\in\mathcal{X}}{\text{maximize}} \; \inf_{F\in\mathcal{D}} \mathbb{E}_F[h(x,\xi)] \; .$$

where $\mathcal{D}$ captures exactly what is known of the distribution

Distribution information in data-driven optimization

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)]

$$\mathcal{S} \;\&\; \{\xi_i\}_{i=1}^M \to \mathcal{D} := \{F \,|\, \exists\, \theta,\ \psi(F) = \theta,\ T_\theta(\{\xi_i\}) \leq \gamma(M)\}$$

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

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

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

$$\mathcal{S} \;\&\; \{\xi_i\}_{i=1}^M \to \mathcal{D} := \{F \,|\, d(F, \hat{F}) \leq \gamma(M)\}$$

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

Gamble A		Urn #2	
		Blue	Red
Urn #1	Blue	180	180
	Red	20	20

Gamble B		Urn #2	
		Blue	Red
Urn #1	Blue	200	200
	Red	0	0

Gamble C		Urn #2	
		Blue	Red
Urn #1	Blue	100	0
	Red	100	0

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:

$$\max_{x\in\{0,1\}^3,\, x_A+x_B+x_C=1} \; \min_{p\in[0,1]^2} (20+160p_1)x_A + (200p_1)x_B + (100p_2)x_C$$

Value of distribution information

- 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

Outline

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 $\mathcal{D} = \cup_{o\in\mathcal{O}}\mathcal{D}(o)$.

- Worst-case value of information:

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

- Best-case value of information

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

- Worst-case regret of not using the information

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

where $x_0 \in \arg\max_{x\in\mathcal{X}} \min_{F\in\mathcal{D}} \mathbb{E}_F[h(x,\xi)]$

Value of distribution information in Two Urns exp.

Gamble A		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	180	180
	Red	20	20

Gamble B		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	200	200
	Red	0	0

Gamble C		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	100	0
	Red	100	0

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

- Based on WC-VDI:

$$\begin{aligned} \text{WC-VDI(Urn\#1)} &= \min_{p_1\in[0,1]} \max_{x\in\mathcal{X}} \min_{p_2\in[0,1]} (20+160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\ &\quad - \max_{x\in\mathcal{X}} \min_{p\in[0,1]^2} (20+160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\ &= \min_{p_1\in[0,1]} \max_{x\in\mathcal{X}} (20+160p_1)x_A + (200p_1)x_B - 20 \\ &= \max_{x\in\mathcal{X}} 20x_A - 20 = 0 \end{aligned}$$

Value of distribution information in Two Urns exp.

Gamble A		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	180	180
	Red	20	20

Gamble B		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	200	200
	Red	0	0

Gamble C		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	100	0
	Red	100	0

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

- Based on WC-VDI:

$$\begin{aligned}\text{WC-VDI(Urn\#1)} &= 0\\ \text{WC-VDI(Urn\#2)} &= \min_{p_2\in[0,1]}\max_{x\in\mathcal{X}}\min_{p_1\in[0,1]}(20+160p_1)x_A+(200p_1)x_B+(100p_2)x_C\\ &\quad-\max_{x\in\mathcal{X}}\min_{p\in[0,1]^2}(20+160p_1)x_A+(200p_1)x_B+(100p_2)x_C\\ &= \min_{p_2\in[0,1]}\max_{x\in\mathcal{X}}20x_A+(100p_2)x_C-20 = \max_{x\in\mathcal{X}}20x_A-20=0\end{aligned}$$

Value of distribution information in Two Urns exp.

Gamble A		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	180	180
	Red	20	20

Gamble B		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	200	200
	Red	0	0

Gamble C		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	100	0
	Red	100	0

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

- Based on WC-VDI:

$$\text{WC-VDI(Urn\#1)} = 20 - 20 = 0 \quad \text{(i.e. confirm no blue in urn \#1.)}$$
$$\text{WC-VDI(Urn\#2)} = 20 - 20 = 0 \quad \text{(i.e. confirm no blue in urn \#2.)}$$

- Conclusion: Distribution information has no value!

Value of distribution information in Two Urns exp.

Gamble A		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	180	180
	Red	20	20

Gamble B		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	200	200
	Red	0	0

Gamble C		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	100	0
	Red	100	0

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

- Based on BC-VDI:

$$\begin{aligned}\text{BC-VDI(Urn\#1)} &= \max_{p_1\in[0,1]} \max_{x\in\mathcal{X}} \min_{p_2\in[0,1]} (20+160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\ &\quad - \max_{x\in\mathcal{X}} \min_{p\in[0,1]^2} (20+160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\ &= \max_{p_1\in[0,1]} \max_{x\in\mathcal{X}} (20+160p_1)x_A + (200p_1)x_B - 20 \\ &= \max_{x\in\mathcal{X}} 180x_A + 200x_B - 20 = 180\end{aligned}$$

Value of distribution information in Two Urns exp.

Gamble A		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	180	180
Urn #1	Red	20	20

Gamble B		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	200	200
Urn #1	Red	0	0

Gamble C		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	100	0
Urn #1	Red	100	0

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

- Based on BC-VDI:

$$\begin{aligned}
\text{BC-VDI(Urn\#1)} &= 180 \\
\text{BC-VDI(Urn\#2)} &= \max_{p_2\in[0,1]} \max_{x\in\mathcal{X}} \min_{p_1\in[0,1]} (20+160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\
&\quad - \max_{x\in\mathcal{X}} \min_{p\in[0,1]^2} (20+160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\
&= \max_{p_2\in[0,1]} \max_{x\in\mathcal{X}} 20x_A + (100p_2)x_C - 20 \\
&= \max_{x\in\mathcal{X}} 20x_A + 100x_B - 20 = 80
\end{aligned}$$

Value of distribution information in Two Urns exp.

Gamble A		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	180	180
	Red	20	20

Gamble B		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	200	200
	Red	0	0

Gamble C		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	100	0
	Red	100	0

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

- Based on BC-VDI:

BC-VDI(Urn#1) $= 200 - 20 = 180$ (i.e. confirm all blue in urn #1.)

BC-VDI(Urn#2) $= 100 - 20 = 80$ (i.e. confirm all blue in urn #2.)

- Conclusion: One should count the balls of Urn #1 !

Value of distribution information in Two Urns exp.

Gamble A		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	180	180
	Red	20	20

Gamble B		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	200	200
	Red	0	0

Gamble C		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	100	0
	Red	100	0

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

- Based on BC-VDI:

BC-VDI(Urn#1) $= 180 - 20 = 160$ (i.e. confirm all blue in urn #1.)

BC-VDI(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 !

Value of distribution information in Two Urns exp.

Gamble A		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	180	180
Urn #1	Red	20	20

Gamble B		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	200	200
Urn #1	Red	0	0

Gamble C		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	100	0
Urn #1	Red	100	0

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

- Based on WCR-VDI:

$$\begin{aligned}\text{WCR-VDI(Urn\#1)} &= \max_{p_1\in[0,1]} \max_{x\in\mathcal{X}} \min_{p_2\in[0,1]} (20+160p_1)x_A + (200p_1)x_B + (100p_2)x_C \\ &\quad - \min_{p_2\in[0,1]} (20+160p_1)\cdot 1 + (200p_1)\cdot 0 + (100p_2)\cdot 0 \\ &= \max_{p_1\in[0,1]} \max_{x\in\mathcal{X}} (20+160p_1)x_A + (200p_1)x_B - (20+160p_1) \\ &= \max_{p_1\in[0,1]} \max_{x\in\mathcal{X}} (40p_1 - 20)x_B = \max_{x\in\mathcal{X}} 20x_B = 20\end{aligned}$$

Value of distribution information in Two Urns exp.

Gamble A		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	180	180
	Red	20	20

Gamble B		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	200	200
	Red	0	0

Gamble C		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	100	0
	Red	100	0

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

- Based on WCR-VDI:

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

Value of distribution information in Two Urns exp.

Gamble A		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	180	180
	Red	20	20

Gamble B		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	200	200
	Red	0	0

Gamble C		Urn #2 Blue	Urn #2 Red
Urn #1	Blue	100	0
	Red	100	0

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 !

Outline

An ordering of VDI measures

In Two Urns experiment, we noticed that

$$\overbrace{0}^{\text{WC-VDI(Urn\#1)}} \leq \overbrace{20}^{\text{WCR-VDI(Urn\#1)}} \leq \overbrace{180}^{\text{BC-VDI(Urn\#1)}}$$

$$\overbrace{0}^{\text{WC-VDI(Urn\#2)}} \leq \overbrace{80}^{\text{WCR-VDI(Urn\#2)}} \leq \overbrace{80}^{\text{BC-VDI(Urn\#2)}}$$

Lemma

It is generally the case that

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

An ordering of VDI measures

Lemma

It is generally the case that

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

Proof:

$$\begin{aligned}
\text{WC-VDI}(\mathcal{O}) &= \min_{o_1\in\mathcal{O}} \max_{x_1} \min_{F\in\mathcal{D}(o_1)} \mathbb{E}_F[h(x_1,\xi)] - \max_{x_2\in\mathcal{X}} \min_{F\in\mathcal{D}} \mathbb{E}_F[h(x_2,\xi)] \\
&= \min_{o_1\in\mathcal{O}} \max_{x_1\in\mathcal{X}} \min_{F\in\mathcal{D}(o_1)} \mathbb{E}_F[h(x_1,\xi)] - \min_{o_2\in\mathcal{O}} \min_{F\in\mathcal{D}(o_2)} \mathbb{E}_F[h(x_0,\xi)] \\
&= \max_{o_2\in\mathcal{O}} \min_{o_1\in\mathcal{O}} \max_{x_1\in\mathcal{X}} \min_{F\in\mathcal{D}(o_1)} \mathbb{E}_F[h(x_1,\xi)] - \min_{F\in\mathcal{D}(o_2)} \mathbb{E}_F[h(x_0,\xi)] \\
&\leq \max_{o_1=o_2\in\mathcal{O}} \max_{x_1\in\mathcal{X}} \min_{F\in\mathcal{D}(o_1)} \mathbb{E}_F[h(x_1,\xi)] - \min_{F\in\mathcal{D}(o_2)} \mathbb{E}_F[h(x_0,\xi)] \\
&= \text{WCR-VDI}(\mathcal{O})
\end{aligned}$$

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:

$$\begin{aligned}
\text{WCR-VDI}(\mathcal{O}) &= \max_{o\in\mathcal{O}}\{\max_{x_1\in\mathcal{X}} \min_{F\in\mathcal{D}(o)} \mathbb{E}_F[h(x_1,\xi)] - \min_{F\in\mathcal{D}(o)} \mathbb{E}_F[h(x_0,\xi)]\} \\
&\leq \max_{o_1\in\mathcal{O}} \max_{o_2\in\mathcal{O}} \max_{x_1\in\mathcal{X}} \min_{F\in\mathcal{D}(o_1)} \mathbb{E}_F[h(x_1,\xi)] - \min_{F\in\mathcal{D}(o_2)} \mathbb{E}_F[h(x_0,\xi)] \\
&= \max_{o_1\in\mathcal{O}} \max_{x_1\in\mathcal{X}} \min_{F\in\mathcal{D}(o_1)} \mathbb{E}_F[h(x_1,\xi)] - \min_{F\in\mathcal{D}} \mathbb{E}_F[h(x_0,\xi)] \\
&= \text{BC-VDI}(\mathcal{O})
\end{aligned}$$

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

$$\begin{aligned}
\text{WC-VDI}(\mathcal{O}) &= \min_{o\in\mathcal{O}} \max_{x_1\in\mathcal{X}} \min_{F\in\mathcal{D}(o)} \mathbb{E}_F[h(x_1,\xi)] - \max_{x_2\in\mathcal{X}} \min_{F\in\mathcal{D}} \mathbb{E}_F[h(x_2,\xi)] \\
&\leq \min_{o\in\mathcal{O}} \min_{F\in\mathcal{D}(o)} \max_{x_1\in\mathcal{X}} \mathbb{E}_F[h(x_1,\xi)] - \max_{x_2\in\mathcal{X}} \min_{F\in\mathcal{D}} \mathbb{E}_F[h(x_2,\xi)] \\
&= \min_{F\in\mathcal{D}} \max_{x_1\in\mathcal{X}} \mathbb{E}_F[h(x_1,\xi)] - \max_{x_2\in\mathcal{X}} \min_{F\in\mathcal{D}} \mathbb{E}_F[h(x_2,\xi)] = 0 \,.
\end{aligned}$$

Some situations where VDI = 0 in the best case

Theorem (Delage et al., 2014)

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

Proof:

$$\begin{aligned} \text{BC-VDI} &= \max_{o \in \mathcal{O}} \max_{x \in \mathcal{X}} \min_{F \in \mathcal{D}(o)} \mathbb{E}_F[h(x, \xi)] - \max_{x \in \mathcal{X}} \min_{F \in \mathcal{D}} \mathbb{E}_F[h(x, \xi)] \\ &= \max_{o \in \mathcal{O}} \max_{x \in \mathcal{X}} \mathbb{E}_F[h(x, \mu)] - \max_{x \in \mathcal{X}} \mathbb{E}_F[h(x, \mu)] = 0 \;, \end{aligned}$$

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

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 ξ, and $\mathcal{D}_{moment}$ is used.

Sketch of proof:

- When the distribution information is perfect,

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

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

$$\begin{aligned} h(x,\xi) := \max_{y\in\mathbb{R}^m} \quad & c^T x + \xi^T y \\ \text{s.t.} \quad & |y_i| \leq x\,,\ \forall y \in \{1,2,\ldots,m\} \\ & a^T y = 0\,. \end{aligned}$$

Tractable bound for worst-case regret

Theorem (Delage et al., 2014)

If the following conditions apply:

1. $\mathcal{D}_{moment}$ *is used with* $\mathcal{S} \subseteq \{\xi \mid \|\xi\|_1 \leq \rho\}$ *and* $\|\mathbb{E}_F[\xi] - \hat{\mu}\|^2_{\hat{\Sigma}^{-1/2}} \leq \gamma_1$
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$(\mathcal{O})$ *can be evaluated*

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

$$\text{s.t.} \quad s \geq \alpha(\rho \boldsymbol{e}_i) - \rho \boldsymbol{e}_i^\mathsf{T} q \,, \; \forall i \in \{1, ..., m\}$$

$$s \geq \alpha(-\rho \boldsymbol{e}_i) + \rho \boldsymbol{e}_i^\mathsf{T} q \,, \; \forall i \in \{1, ..., m\} \,,$$

where $\alpha(\xi) = \max_{x \in \mathcal{X}} h(x,\xi) - (c^T \bar{y}_0 + \xi^T C \bar{y}_0)$ *for any* $\bar{y}_0 \in \mathcal{Y}(x_0)$.

Outline

Value of distribution info for an airline company

- 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 ?

Mathematical formulation for fleet mix optimization

The fleet composition problem is a stochastic mixed integer LP

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

Fleet mix → x

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

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

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

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

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

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:

Test cases	WCR-VDI(o)
#1	$\leq 6\%$
#2	$\leq 1\%$
#3	$\leq 7\%$

Conclusions:

- It's wasteful in these problems to invest more than 7% of profits in acquisition of distribution information

Outline

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

References I

A. Ben-Tal, D. den Hertog, A. Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. working draft, 2015.

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

E. Delage, S. Arroyo, and Y. Ye. The value of stochastic modeling in two-stage stochastic programs with cost uncertainty. *Operations Research*, 62(6):1377–1393, 2014.

E. Erdogan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming Series B*, 107:36–61, 2006.

P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. Technical report, 2015.

H. Scarf. A min-max solution of an inventory problem. *Studies in The Mathematical Theory of Inventory and Production*, pages 201–209, 1958.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

Questions & Comments ...

... Thank you!