The Value of Stochastic Modeling in Two-Stage Stochastic Programs

Erick Delage, HEC Montréal
Sharon Arroyo, The Boeing Cie.
Yinyu Ye, Stanford University

Tuesday, October 8th, 2013

Two-Stage Stochastic Programming

Let's consider the stochastic programming problem:

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

- $\boldsymbol{x}$ is a vector of decision variables in $\mathbb{R}^n$
- $\boldsymbol{\xi}$ is a vector of uncertain parameters in $\mathbb{R}^d$

Two-Stage Stochastic Programming

Let's consider the stochastic programming problem:

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

- $\boldsymbol{x}$ is a vector of decision variables in $\mathbb{R}^n$
- $\boldsymbol{\xi}$ is a vector of uncertain parameters in $\mathbb{R}^d$
- The profit function $h(\boldsymbol{x}, \boldsymbol{\xi})$ is the maximum of a linear program with uncertainty limited to objective

$$\begin{aligned} h(\boldsymbol{x}, \boldsymbol{\xi}) \;:=\; \max_{\boldsymbol{y}}. \quad & \boldsymbol{c}_1^{\mathsf{T}}\boldsymbol{x} + \boldsymbol{\xi}^{\mathsf{T}} C_2 \boldsymbol{y} \\ \text{s.t.} \quad & A\boldsymbol{x} + B\boldsymbol{y} \leq \boldsymbol{b} \end{aligned}$$

Two-Stage Stochastic Programming

Let's consider the stochastic programming problem:

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

- $\boldsymbol{x}$ is a vector of decision variables in $\mathbb{R}^n$
- $\boldsymbol{\xi}$ is a vector of uncertain parameters in $\mathbb{R}^d$
- The profit function $h(\boldsymbol{x},\boldsymbol{\xi})$ is the maximum of a linear program with uncertainty limited to objective

$$h(\boldsymbol{x},\boldsymbol{\xi}) := \begin{array}{ll} \max\limits_{\boldsymbol{y}}. & \boldsymbol{c}_1^\mathsf{T}\boldsymbol{x} + \boldsymbol{\xi}^\mathsf{T} C_2 \boldsymbol{y} \\ \text{s.t.} & A\boldsymbol{x} + B\boldsymbol{y} \leq \boldsymbol{b} \end{array}$$

- To find an optimal solution, one must develop a stochastic model and solve the associated stochastic program

Difficulty of Developing a Stochastic Model

- Developing an accurate stochastic model requires heavy engineering efforts and might even be impossible:
  - Expecting that a scenario might occur does not determine its probability of occurring
  - Unexpected events (e.g., economic crisis) might occur
  - The future might actually not behave like the past

Difficulty of Developing a Stochastic Model

- Developing an accurate stochastic model requires heavy engineering efforts and might even be impossible:
  - Expecting that a scenario might occur does not determine its probability of occurring
  - Unexpected events (e.g., economic crisis) might occur
  - The future might actually not behave like the past

- What if, after all this work, we realize that the solution only marginally improves performance?

Difficulty of Developing a Stochastic Model

- Developing an accurate stochastic model requires heavy engineering efforts and might even be impossible:
  - Expecting that a scenario might occur does not determine its probability of occurring
  - Unexpected events (e.g., economic crisis) might occur
  - The future might actually not behave like the past

- What if, after all this work, we realize that the solution only marginally improves performance?
- What if, after implementing the SP solution, we realize that our choice of distribution was wrong?

A few data-driven approaches

In practice, we often have loads of historical data to inform our decision. We can consider a number of data-driven approaches:

A few data-driven approaches

In practice, we often have loads of historical data to inform our decision. We can consider a number of data-driven approaches:

- Mean value problem: estimate $\mathbb{E}[\xi]$ and solve

$$\text{(MVP)} \qquad \underset{x\in\mathcal{X}}{\text{maximize}} \;\; h(x, \mathbb{E}[\xi]) \; .$$

A few data-driven approaches

In practice, we often have loads of historical data to inform our decision. We can consider a number of data-driven approaches:

- Mean value problem: estimate $\mathbb{E}[\xi]$ and solve

$$\text{(MVP)} \qquad \underset{x\in\mathcal{X}}{\text{maximize}} \;\; h(x, \mathbb{E}[\xi]) \;\; .$$

- Empirical Average Approximation: solve

$$\text{(EAA)} \qquad \underset{x\in\mathcal{X}}{\text{maximize}} \;\; \frac{1}{M}\sum_i h(x, \xi_i) \;\; .$$

A few data-driven approaches

In practice, we often have loads of historical data to inform our decision. We can consider a number of data-driven approaches:

- Mean value problem: estimate $\mathbb{E}[\xi]$ and solve

$$\text{(MVP)} \qquad \underset{x \in \mathcal{X}}{\text{maximize}} \;\; h(x, \mathbb{E}[\xi]) \;\; .$$

- Empirical Average Approximation: solve

$$\text{(EAA)} \qquad \underset{x \in \mathcal{X}}{\text{maximize}} \;\; \frac{1}{M} \sum_{i} h(x, \xi_i) \;\; .$$

- Distributionally robust problem: use data to characterize information about the moments $\mu, \Sigma, etc.$ and solve:

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

A few data-driven approaches

In practice, we often have loads of historical data to inform our decision. We can consider a number of data-driven approaches:

- Expected value problem
- Empirical Average Approximation
- Distributionally robust problem

How can we find out if we would achieve more with a stochastic model

A few data-driven approaches

In practice, we often have loads of historical data to inform our decision. We can consider a number of data-driven approaches:

- Expected value problem
- Empirical Average Approximation
- Distributionally robust problem

How can we find out if we would achieve more with a stochastic model without developing the stochastic model?

Outline

Outline

Distributionally Robust Optimization

Use available information to define a set $\mathcal{D}$, such that $F \in \mathcal{D}$, then consider the distributionally robust stochastic program:

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

Distributionally Robust Optimization

Use available information to define a set $\mathcal{D}$, such that $F \in \mathcal{D}$, then consider the distributionally robust stochastic program:

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

- Introduced by H. Scarf in 1958
- Generalizes many forms of optimization models

  E.g.: stochastic programming, robust optimization, deterministic optimization
- Many instances have been shown to be easier to solve than the associated SP

  [Calafiore et al. (2006), Delage et al. (2010), Chen et al. (2010)]

Finite sample guarantees for a DRSP

Theorem (Delage & Ye, 2010)

If the data is i.i.d., then the solution to the DRSP under the uncertainty set

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

with $\gamma_1 = O\left(\frac{\log(1/\delta)}{M}\right)$ *and* $\gamma_2 = O\left(\sqrt{\frac{\log(1/\delta)}{M}}\right)$*, achieves an expected performance that is guaranteed, with prob. greater than* $1-\delta$*, to be better than the optimized value of the DRSP problem.*

Value of MVP solution under Bounded Moments

Theorem (Delage, Arroyo & Ye, 2013)

Given that the stochastic program is risk neutral, the solution to the MVP is optimal with respect to

$$\underset{x\in\mathcal{X}}{\text{maximize}} \quad \inf_{F\in\mathcal{D}(\mathcal{S},\hat{\boldsymbol{\mu}},\hat{\Sigma})} \mathbb{E}_F[h(\boldsymbol{x},\boldsymbol{\xi})] \ ,$$

where

$$\mathcal{D}(\mathcal{S},\hat{\boldsymbol{\mu}},\hat{\Sigma}) = \left\{ F \left| \begin{array}{l} \mathbb{P}(\boldsymbol{\xi}\in\mathcal{S}) = 1 \\ \|\mathbb{E}\left[\boldsymbol{\xi}\right] - \hat{\mu}\|^2_{\hat{\Sigma}^{-1/2}} \leq 0 \\ \mathbb{E}\left[(\boldsymbol{\xi}-\hat{\boldsymbol{\mu}})(\boldsymbol{\xi}-\hat{\boldsymbol{\mu}})^{\mathsf{T}}\right] \preceq (1+\gamma_2)\hat{\Sigma} \end{array} \right. \right\}$$

Finite sample guarantees for Robust MVP

Corollary

If the data is i.i.d., then the solution to the Robust MVP

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{maximize}} \quad \min_{\boldsymbol{\mu}:\|\hat{\Sigma}^{-1/2}(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})\|^2\leq\gamma_1} h(\boldsymbol{x},\boldsymbol{\mu}) \ .$$

with $\gamma_1 = O\left(\frac{\log(1/\delta)}{M}\right)$ *achieves an expected performance that is guaranteed, with prob. greater than* $1-\delta$*, to be better than the optimized value of the Robust MVP problem.*

Inferring structure from data

In practice, we often know something about the structure of $\boldsymbol{\xi}$

- Factor model: $\boldsymbol{\xi} = \boldsymbol{c} + \boldsymbol{A}\boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \in \mathbb{R}^{d'}$, $d' << d$
- Autoregressive-moving-average (ARMA) model:

$$\xi_t = c + \sum_{j=1}^{p} \psi_j \xi_{t-i} + +\varepsilon_t \sum_{j=1}^{q} \theta_i \varepsilon_{t-i} \text{ with } \varepsilon_t \text{ i.i.d.}$$

- Autoregressive Conditional Heteroskedasticity (ARCH)

$$\xi_t = c_t + \sigma_t \varepsilon_t, \quad \sigma_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j (\sigma_{t-j}\varepsilon_{t-j})^2 \text{ with } \varepsilon_t \text{ i.i.d.}$$

Do we need to make distribution assumptions to calibrate these models ?

Generalized method of moments [A.R. Hall (2005)]

Suppose that the structure is $\xi_t = \varepsilon_t + \theta\varepsilon_{t-1}$, with ε_t i.i.d. with mean μ and σ

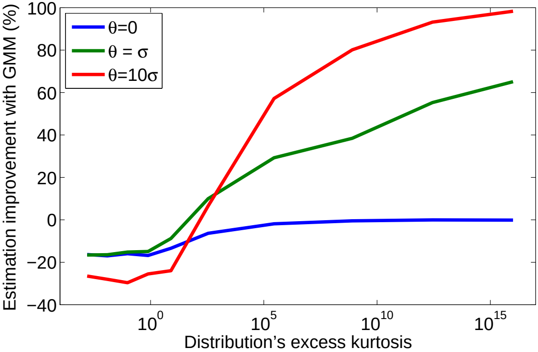
- Regardless of the distribution of $\boldsymbol{\varepsilon}$, we know that

$$\begin{aligned}
\mathbb{E}[\xi_t] &= (1+\theta)\mu \\
\mathbb{E}[\xi_t\xi_{t-1}] &= \mathbb{E}[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] \\
&= \mu^2 + \theta\mu^2 + \theta(\mu^2+\sigma^2) + \theta^2\mu^2 = (1+\theta)^2\mu^2 + \theta\sigma^2 \\
\mathbb{E}[\xi_t^2] &= (1+\theta)^2\mu^2 + (1+\theta^2)\sigma^2
\end{aligned}$$

- Use empirical moments to fit the parameters (θ, μ, σ)
- Retrieve the moments for $\boldsymbol{\xi}$

Quality of GMM estimation

Empirical evaluation of quality of covariance estimation using GMM versus Gaussian likelihood maximization when true distribution is log-normal



Outline

What is the Value of Stochastic Modeling?

Consider the following steps:

1. Construct $\mathcal{D}$ such that $F \in \mathcal{D}$ with high confidence
2. Find candidate solution using data-driven approach
3. Is it worth developing a stochastic model?
   (a) If yes, then develop a model & solve SP
   (b) Otherwise, implement candidate solution

What is the Value of Stochastic Modeling?

Consider the following steps:

1. Construct $\mathcal{D}$ such that $F \in \mathcal{D}$ with high confidence
2. Find candidate solution using data-driven approach
3. Is it worth developing a stochastic model?
   (a) If yes, then develop a model & solve SP
   (b) Otherwise, implement candidate solution

Worst-case regret of a candidate solution gives an optimistic estimate of the value of obtaining perfect information about F.

$$\mathcal{R}(x_1) \;:=\; \sup_{F \in \mathcal{D}} \left\{ \max_{x_2} \mathbb{E}_F[h(x_2, \xi)] - \mathbb{E}_F[h(x_1, \xi)] \right\}$$

What is the Value of Stochastic Modeling?

Consider the following steps:

1. Construct $\mathcal{D}$ such that $F \in \mathcal{D}$ with high confidence
2. Find candidate solution using data-driven approach
3. Is it worth developing a stochastic model?
   (a) If yes, then develop a model & solve SP
   (b) Otherwise, implement candidate solution

Worst-case regret of a candidate solution gives an optimistic estimate of the value of obtaining perfect information about F.

$$\mathcal{R}(\boldsymbol{x}_1) \;:=\; \sup_{F\in\mathcal{D}} \left\{ \max_{\boldsymbol{x}_2} \mathbb{E}_F[h(\boldsymbol{x}_2,\boldsymbol{\xi})] - \mathbb{E}_F[h(\boldsymbol{x}_1,\boldsymbol{\xi})] \right\}$$

Theorem (Delage, Arroyo & Ye, 2013)

Evaluating the worst-case regret $\mathcal{R}(\boldsymbol{x}_1)$ exactly is NP-hard in general.

Bounding the Worst-case Regret

Theorem (Delage, Arroyo & Ye, 2013)

If $\mathcal{S} \subseteq \{\boldsymbol{\xi} \mid \|\boldsymbol{\xi}\|_1 \leq \rho\}$ *and* $\|\mathbb{E}_F[\boldsymbol{\xi}] - \hat{\boldsymbol{\mu}}\|^2_{\hat{\Sigma}^{-1/2}} \leq \gamma_1$*, then an upper bound can be evaluated*

$$
\begin{aligned}
\mathcal{UB}(\boldsymbol{x}_1, \bar{\boldsymbol{y}}_1) \;:=\; \min_{s,\boldsymbol{q}} \quad & s + \hat{\boldsymbol{\mu}}^{\mathsf{T}}\boldsymbol{q} + \sqrt{\gamma_1}\|\hat{\Sigma}^{1/2}\boldsymbol{q}\| \\
\text{s.t.} \quad & s \geq \alpha(\rho \boldsymbol{e}_i) - \rho \boldsymbol{e}_i^{\mathsf{T}}\boldsymbol{q} \;, \forall i \in \{1, ..., d\} \\
& s \geq \alpha(-\rho \boldsymbol{e}_i) + \rho \boldsymbol{e}_i^{\mathsf{T}}\boldsymbol{q} \;, \forall i \in \{1, ..., d\} \,,
\end{aligned}
$$

where $\alpha(\boldsymbol{\xi}) = \max_{\boldsymbol{x}_2} h(\boldsymbol{x}_2, \boldsymbol{\xi}) - h(\boldsymbol{x}_1, \boldsymbol{\xi}; \bar{\boldsymbol{y}}_1)$.

Outline

Value of Stochastic Modeling for an Airline Company

- 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: passenger demand, fuel prices, etc.
- Yet, many airline companies sign these contracts based on a single scenario of what the future may be.

Value of Stochastic Modeling for an Airline Company

- 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:
    passenger demand, fuel prices, etc.
- Yet, many airline companies sign these contracts based on a single scenario of what the future may be.
- Now we know that since little is known about these uncertain factors, using the data-driven forecast of expected value of parameters can be considered a robust approach

Value of Stochastic Modeling for an Airline Company

- 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: passenger demand, fuel prices, etc.
- Yet, many airline companies sign these contracts based on a single scenario of what the future may be.
- Now we know that since little is known about these uncertain factors, using the data-driven forecast of expected value of parameters can be considered a robust approach
- Can we do better by developing a stochastic model ?

Mathematical Formulation for Fleet Mix Optimization

The fleet composition problem is a stochastic mixed integer LP

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

Fleet mix → $\boldsymbol{x}$

Mathematical Formulation for Fleet Mix Optimization

The fleet composition problem is a stochastic mixed integer LP

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

Fleet mix → $\boldsymbol{x}$

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

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

s.t.

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

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

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

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, $\sigma_{\tilde{p}_i}/\mu_{\tilde{p}_i} \in [2\%, 58\%]$

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	Worst-case regret for MVP solution
#1	$\leq 6\%$
#2	$\leq 1\%$
#3	$\leq 7\%$

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	Worst-case regret for MVP solution
#1	$\leq 6\%$
#2	$\leq 1\%$
#3	$\leq 7\%$

Conclusions:

- It's wasteful to invest more than 7% of profits in stochastic modeling

Outline

Conclusion

- A lot can be done with data before developing a stochastic model
  - A distributionally robust model formulated using the data can provide useful guarantees
  - In some circumstances, the MVP model provides a distributionally robust solution
  - It is possible to calibrate a structural model using GMM

Conclusion

- A lot can be done with data before developing a stochastic model
  - A distributionally robust model formulated using the data can provide useful guarantees
  - In some circumstances, the MVP model provides a distributionally robust solution
  - It is possible to calibrate a structural model using GMM
- One can estimate how much might be gained with a stochastic model

Conclusion

- A lot can be done with data before developing a stochastic model
  - A distributionally robust model formulated using the data can provide useful guarantees
  - In some circumstances, the MVP model provides a distributionally robust solution
  - It is possible to calibrate a structural model using GMM
- One can estimate how much might be gained with a stochastic model
- In some cases, using the data itself might be good enough

Bibliography

Calafiore, G., L. El Ghaoui. 2006. On distributionally robust chance-constrained linear programs. *Optimization Theory and Applications* **130**(1) 1–22.

Chen, W., M. Sim, J. Sun, C.-P. Teo. 2010. From CVaR to uncertainty set: Implications in joint chance constrained optimization. *Operations Research* **58** 470–485.

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.

Hall, A. R. 2005. *Generalized Method of Moments*. 1st ed. Oxford University Press.

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

Questions & Comments ...

... Thank you!