Distributionally Robust Optimization under Moment Uncertainty with Application to Data-Driven Problems

Erick Delage, edelage@stanford.edu
Yinyu Ye, yinyu-ye@stanford.edu
Stanford University

October 2008
Uncertainty in Optimization

Consider an optimization problem:

\[ \min_{x \in \mathcal{X}} h(x, \xi) \]

- \( x \) is the variable
- \( \xi \) is a vector of parameters

We often cannot assume to know \( \xi \) exactly and need ways to account for uncertainty:

- deterministic \( \xi \) estimated from noisy measurements
- \( \xi \) is actually a random parameter
Stochastic Programming

Assume $\xi \sim f_\xi$, solve:

$$\minimize_{x \in \mathcal{X}} \mathbb{E}_{f_\xi}[h(x, \xi)]$$

- Somewhat tractable (sample average approx.)
- Science goes into formulating distribution $f_\xi$

What if I don’t know $f_\xi$?

- Access to limited information about $f_\xi$ (e.g. samples)
- Future might not be distributed like the past
- Solution might be sensitive to my choice of $f_\xi$
Distributionally Robust Optimization

Describe uncertainty in $f_{\xi}$ through a set $D$ and consider worst case distribution:

$$(DRSP) \quad \text{minimize} \quad \max_{x \in \mathcal{X}} \quad \text{max}_{f_{\xi} \in D} \quad \mathbb{E}_{f_{\xi}}[h(x, \xi)] \, ,$$

In this talk, we consider that one can often describe his uncertainty in $f_{\xi}$ through:

$$D(\gamma) = \left\{ f_{\xi} \mid \begin{array}{l}
\mathbb{P}(\xi \in S) = 1 \\
(\mathbb{E}[\xi] - \hat{\mu})^T \hat{\Sigma}^{-1} (\mathbb{E}[\xi] - \hat{\mu}) \leq \gamma_1 \\
\mathbb{E}[(\xi - \hat{\mu})(\xi - \hat{\mu})^T] \preceq (1 + \gamma_2)\hat{\Sigma}
\end{array} \right\} \, ,$$

for some $\gamma_1, \gamma_2 \geq 0$. 

Related Work

- DRSP is related to the moment problem
  [Bertsimas & Popescu, 2005]

- If only support is known, then reduces to robust optimization
  [Ben-Tal & Nemirovski, 1998]

- If mean & variance are known, special cases have been studied: e.g., newsvendor, portfolio selection, linear chance constraints, etc.
  [Scarf, 1958; Popescu, 2007; Calafiore & El Ghaoui, 2006]

- Prior to this work, little was known about how practical the DRSP is with $\mathcal{D}(\gamma)$
Outline

- Solving the Dist. Robust Stochastic Program
  - Conditions on $h(x, \xi)$ for existence of a tractable solution

- Using the DRSP in data-driven problems
  - Modeling the DRSP to get high confidence solutions

- Dist. Robust Portfolio Optimization
  - Empirical evaluation with real stock market data
Conditions on $h(x, \xi)$

Let $h(x, \xi) = \max_{k \in \{1, 2, \ldots, K\}} h_k(x, \xi)$ such that for all $k$:

- $h_k(x, \xi)$ is convex in $x$
- $h_k(x, \xi)$ is concave in $\xi$
- $h_k(x, \xi)$’s value and gradient are easily obtained

Examples:

- $h(x, \xi) = -u(x^T \xi)$ with $u(\cdot)$ piecewise linear concave
- Many $h(x, \xi)$ for which simpler robust form is tractable: i.e., minimize$_{x \in X}$ max$_{\xi \in S}$ $h(x, \xi)$.
Theorem 1. : *In the case that* $h(x, \xi)$ *satisfies our conditions, the following distributionally robust problem:

$$\min_{x \in \mathcal{X}} \max_{f_\xi \in \mathcal{D}} \mathbb{E}_{f_\xi}[h(x, \xi)]$$

1. *is NP-hard for general* $\mathcal{D}$ *(Bertsimas 2005)*
2. *can be solved in polynomial time for* $\mathcal{D}(\gamma)$
3. *stands with* $\mathcal{D}(0)$ *as a valuable relaxation of some NP-hard forms*
Solution Method

The following problem can be solved in polynomial time:

\[
\begin{align*}
\text{minimize} \quad & \max_{f_\xi \in \mathcal{D}(\gamma)} \mathbb{E}_{f_\xi} [h(x, \xi)] \\
\text{subject to} \quad & x \in \mathcal{X}
\end{align*}
\]

After replacing \( \max_{f_\xi \in \mathcal{D}(\gamma)} \mathbb{E}_{f_\xi} [h(x, \xi)] \) with its dual form,

\[
\begin{align*}
\text{minimize} \quad & r + \left( \gamma_2 \hat{\Sigma} + \hat{\mu} \hat{\mu}^T \right) \cdot Q + \hat{\mu}^T q + \sqrt{\gamma_1} \| \hat{\Sigma}^{1/2} (q + 2Q\hat{\mu}) \| \\
\text{subject to} \quad & r \geq \max_{\xi \in \mathcal{S}} \left( h_k(x, \xi) - \xi^T Q \xi - \xi^T q \right) , \quad \forall \ k \in \{1, \ldots, K\} \\
& Q \succeq 0 \quad , \quad x \in \mathcal{X}
\end{align*}
\]

and applying the ellipsoid method.
Outline

- Solving the Dist. Robust Stochastic Program
  - Conditions on $h(x, \xi)$ for existence of a tractable solution

- Using the DRSP in data-driven problems
  - Modeling the DRSP to get high confidence solutions

- Dist. Robust Portfolio Optimization
  - Empirical evaluation with real stock market data
Confidence region for $f_\xi$

Given that:

- I have $M$ i.i.d. samples drawn from $f_\xi$
- I let $\hat{\mu}$ and $\hat{\Sigma}$ be empirical estimates
- I know that $\xi$ lies in a ball of radius $R$

Then, based on concentration of $f_\xi$ [McDiarmid, 1998], for some $\bar{\gamma}_1 = O \left( \frac{R^2}{M} \log(1/\delta) \right)$ and $\bar{\gamma}_2 = O \left( \frac{R^2}{\sqrt{M}} \sqrt{\log(1/\delta)} \right)$, we show that:

$$\mathbb{P}(f_\xi \in \mathcal{D}(\bar{\gamma})) \geq 1 - \delta .$$
Solutions to Data-driven Problems

Given a set of samples \( \{\xi_i\}_{i=1}^M \), build the distributional set \( \mathcal{D}(\gamma) \) then, solve:

\[
\minimize_{x \in \mathcal{X}} \max_{f_{\xi} \in \mathcal{D}(\gamma)} \mathbb{E}_{f_{\xi}}[h(x, \xi)] ,
\]

Conclude that, with high probability, the solution has best worst case expected performance over a set of distributions that contains \( f_{\xi} \).
Outline

- Solving the Dist. Robust Stochastic Program
  - Conditions on $h(x, \xi)$ for existence of a tractable solution

- Using the DRSP in data-driven problems
  - Modeling the DRSP to get high confidence solutions

- Dist. Robust Portfolio Optimization
  - Empirical evaluation with real stock market data
Portfolio Optimization

Stochastic Program form:

$$\maximize_{x \in \mathcal{X}} \mathbb{E}_{f_\xi}[u(\xi^T x)]$$

- $x^j$ is how much do I invest in stock $j$
- Stock $j$ returns in one day $\xi^j$ for each dollar invested
- $u(\cdot)$ is a piecewise linear concave utility function

Distributionally Robust Portfolio Optimization form:

$$(DRPO) \quad \maximize_{x \in \mathcal{X}} \min_{f_\xi \in \mathcal{D}(\gamma)} \mathbb{E}_{f_\xi}[u(\xi^T x)]$$
Experiments with Historical Data

30 stocks were tracked over horizon (1992-2007)
Experiments with Historical Data

An experiment consists of trading 4 stocks over (2001-07).

- Use (1992-2001) to choose $\gamma_1$ and $\gamma_2$
- Update portfolio on daily basis
- Estimate $\hat{\mu}$ and $\hat{\Sigma}$ based on a 30 days period
- DRPO with $\mathcal{D}(\gamma)$ is compared to:
  - DRPO without moment uncertainty
  - Stochastic Program using empirical distribution over last 30 days
Experimental Results

Comparison of wealth evolution in 300 experiments conducted over the years 2001-2007. For each model, the periodical 10% and 90% percentiles of wealth are indicated.
Summary

- Presented a DRSP model that accounts for uncertainty in the parameters of an optimization model.

- Showed that this DRSP model is a natural one to use in data-driven problems.

- Provided insights on how to formulate/recognize problems that are tractable.

- Justified empirically the need to account for both distribution & moment uncertainty.

- We encourage using a distributionally robust criterion as an objective or a constraint; hence, account for risks related to model ambiguity.
Questions & Comments ...

... Thank you!
Experimental Results II

In finer details:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DRPO model</td>
<td>0.944</td>
<td>0.846</td>
<td>1.1017</td>
<td>1.025</td>
</tr>
<tr>
<td>Popescu’s DRPO model</td>
<td>0.700</td>
<td>0.334</td>
<td>1.047</td>
<td>0.9364</td>
</tr>
<tr>
<td>SP model</td>
<td>0.908</td>
<td>0.694</td>
<td>1.045</td>
<td>0.923</td>
</tr>
</tbody>
</table>

79% of the time, our DRPO outperformed both models

On average accounting for moment uncertainty led to a relative gain of 1.67