Two-Stage Stochastic Linear Programs with Incomplete Information on Uncertainty

J. Ang, F. Meng, and J. Sun
National University of Singapore
ICSP, Bergamo, July 9, 2013
jsun@nus.edu.sg
1. Two-Stage Stochastic Optimization with Fixed Recourse (2-SO)

\[
\min_{x \in X} \left\{ c'x + E_P[Q(x, \tilde{z})] \right\}
\]

where the apostrophe ('') stands for the transpose and

\[
Q(x, \tilde{z}) = \min \quad d(\tilde{z})'y(\tilde{z})
\]

s. t. \quad A(\tilde{z})x + Dy(\tilde{z}) = b(\tilde{z}),

\[
y(\tilde{z}) \geq 0,
\]

where \(x \in \mathbb{R}^n\) is the vector of first-stage decision variables subject to a feasible region \(X \subseteq \mathbb{R}^n\)
Problems:

1. Hard or impossible to know the distribution of $\tilde{z}$. Even if the distribution is known, it is often extremely hard to compute the expected value $\mathbb{E}_P(Q(x, \tilde{z}))$.

2. The second stage is risk-neutral, therefore not suitable for some applications.

It is reasonable to consider a stochastic optimization model in which $\tilde{z}$ is structured and only certain partial information on $\tilde{z}$, such as the first and second moments, is known.
Model 1: A worst-case model

We consider the worst case of the recourse value \( \mathbb{E}_P[Q(x, \tilde{z})] \) among all \( P \in \mathcal{F} \) and select \( x \) in such a way that the aggregated worst case cost is minimized. In other words, we are concerned with the following “worst-case version” of problem.

\[
\min_{x \in X} \left\{ c'x + \max_{P \in \mathcal{F}} \mathbb{E}_P[Q(x, \tilde{z})] \right\},
\]

(2)

where \( \mathcal{F} \) is a set of probability measures on which we only have limited knowledge, e.g.,

\[
\mathcal{F} = \{ P : \mathbb{E}_P(\tilde{z}) = \mu, \mathbb{E}_P(\tilde{z}_i^2) \leq \eta_i^2, i = 1, ..., m \}.
\]

This is similar to the idea of distributionally robust SP models (Scarf (1958), Landau (1987), Dupacova (1987), Kall and Wallace (1994), and Delage and Ye (2010), etc.)
Model 2: Using risk measures in second-stage

A generalization of the distributionally robust SP is as follows.

Given a probability space \((\Omega, \Sigma, \mathcal{P}_0)\), we say that \(X : \Omega \to \mathbb{R}\) is a random variable if it is \(\Sigma\)-measurable, that is, \(\{\omega : X(\omega) \leq a\} \in \Sigma\) for any \(a \in \mathbb{R}\).

A risk measure \(\mathcal{R}\) is a functional from \(L^2\) to \((-\infty, +\infty]\). It represents “the risk of loss” where \(X\) represents “the real amount of loss”.

A risk measure $\mathcal{R}$ is \textit{coherent in the basic sense} ("coherent" for short) if it satisfies the following five axioms.

\textbf{(A1)} $\mathcal{R}(C) = C$ for all constant $C$,

\textbf{(A2)} $\mathcal{R}((1 - \lambda)X + \lambda X') \leq (1 - \lambda)\mathcal{R}(X) + \lambda\mathcal{R}(X')$ for $\lambda \in [0, 1]$ ("convexity"),

\textbf{(A3)} $\mathcal{R}(X) \leq \mathcal{R}(X')$ if $X \leq X'$ almost everywhere ("monotonicity"),

\textbf{(A4)} $\mathcal{R}(X) \leq 0$ when $\|X^k - X\|_2 \to 0$ with $\mathcal{R}(X^k) \leq 0$ ("closedness"),

\textbf{(A5)} $\mathcal{R}(\lambda X) = \lambda\mathcal{R}(X)$ for $\lambda > 0$ ("positive homogeneity").

A coherent risk measure is \textit{risk-averse} if $\mathcal{R}(X) > \mathbb{E}(X)$ for all nonconstant $X$. 
Examples.

1. $\mathcal{R}(X) = \mathbb{E}_{\mathbb{P}_0}(X) \equiv \mathbb{E}(X)$ (expectation)

2. $\mathcal{R}(X) = \text{ess.sup}_{\omega \in \Omega} X(\omega)$ (worst case)

3. The conditional value at risk

$$\text{CVaR}_\alpha(X) = \min_{\beta} \left\{ \beta + \frac{1}{1 - \alpha} \mathbb{E}(X - \beta^+) \right\}$$

Dual representation of coherent risk measures

Let the probability measure of $X$ be $\mathbb{P}_0$. Consider another probability measure $\mathbb{P}$ on $(\Omega, \Sigma)$, we say that a probability measure $\mathbb{P}$ is \textit{absolutely continuous} with respect to the base probability measure $\mathbb{P}_0$ (denoted by $\mathbb{P} \ll \mathbb{P}_0$) if $\mathbb{P}_0(A) = 0$ implies $\mathbb{P}(A) = 0$ for any measurable set $A \in \Sigma$.

Let

$$\mathcal{M} := \left\{ \mathbb{P} : \mathbb{P} \ll \mathbb{P}_0, 0 \leq \frac{d\mathbb{P}}{d\mathbb{P}_0} \in L^2, \mathbb{P}(\Omega) = 1 \right\}, \mathcal{P} = \left\{ Q : Q = \frac{d\mathbb{P}}{d\mathbb{P}_0}, \mathbb{P} \in \mathcal{M} \right\}.$$
**Dual Representations of Risk Measures**

$\mathcal{R}$ is a coherent risk measure in the basic sense if and only if there exists a (unique) convex and closed set of densities $Q$, such that for any $X$,

$$\mathcal{R}(X) = \sup_{Q \in Q} \mathbb{E}(XQ) = \sup_{P \in \mathcal{F}} \mathbb{E}_P(X),$$

(3)

where $\mathcal{F} = \{P \in \mathcal{M} : \frac{dP}{dP_0} \in Q\}$.

This is called the "**dual representation of risk measure**", and $Q$ is called the "**risk envelope**" of the risk measure $\mathcal{R}$. 
2. Some tractable cases

We found the new “two-stage risk-measure optimization” model

\[
\min_{x \in X} \left\{ c'x + R[Q(x, \tilde{z})] \right\}.
\]

attractive both in computation and in applications. We studied several cases, most of which can be reduced to second-order cone programs or semidefinite programs.

**A. The risk-neutral case** \( R(X) = E(X) \) with moment constraints \( \mathcal{F} = \{ P : E_P(\tilde{z}) = \mu, E_P(\tilde{z}_i^2) \leq \eta_i^2, i = 1, \ldots, m \} \)

**A1** If stochasticity happens only on the right hand side, i.e. \( b(\tilde{z}) = \tilde{z} \) and if the extreme points of \( \{ u : D'u \leq d \} \) are available, then

\[
\text{Problem (4) } \iff \text{ A second-order cone program of size } O(n + m)
\]
(A2) If $A(\tilde{z}), b(\tilde{z}), d(\tilde{z})$, and $y(\tilde{z})$ satisfy the linear decision rule, namely,

$$
y(\tilde{z}) = y^0 + \sum_{j=1}^{m} \tilde{z}_j y_j^j, \quad b(\tilde{z}) = b^0 + \sum_{j=1}^{m} \tilde{z}_j b_j^j, \quad d(\tilde{z}) = d^0 + \sum_{j=1}^{m} \tilde{z}_j d_j^j, \quad \text{and}
$$

$$
A(\tilde{z}) = A_0 + \sum_{j=1}^{m} \tilde{z}_j A_j,
$$

then

$$
\text{Problem (4)} \iff \text{A second-order cone program of size } O(n + m)
$$
B. The risk-averse case \( \mathcal{R}(X) = \lambda \mathbb{E}(X) + \mu \text{CVaR}(X) \) with moment constraints

\[
\mathcal{F} = \mathcal{F}(\Omega, \mu, \sigma, H, \gamma, \gamma_0) = \left\{ \mathbb{P}(\Omega) = 1, \mathbb{P}: \left| \mathbb{E}_\mathbb{P}[\tilde{z}_i - \mu_i] \right| \leq \sigma_i \gamma_i, i = 1, \ldots, m, \mathbb{E}_\mathbb{P}[\tilde{z}\tilde{z}'] \preceq \gamma_0 H + \mu\mu' \right\},
\]

\((B1)\) If stochasticity happens only on the second-stage objective function, i.e. \( d(\tilde{z}) = \tilde{z} \), then

Problem (4) \iff A semidefinite program of size \( O(n + m) \)

\((B2)\) If stochasticity happens also in constraints, then the problem is generally NP-hard.

\((B3)\) If stochasticity happens also in constraints, then under linear decision rule and assume the extreme points of \( \{ u : D' u \leq d \} \) are available, then

Problem (4) \iff A very large semidefinite program
3. A Numerical Example

**Example.** A company manager is considering the amount of steel to purchase (at $58/lb) for producing wrenches and pliers in next month. The manufacturing process involves molding the tools on a molding machine and then assembling the tools on an assembly machine. Here is the technical data.

<table>
<thead>
<tr>
<th></th>
<th>Wrench</th>
<th>Plier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steel (lbs.)</td>
<td>1.5</td>
<td>1</td>
</tr>
<tr>
<td>Molding Machine (hours)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Assembly Machine (hours)</td>
<td>.3</td>
<td>.5</td>
</tr>
<tr>
<td>Contribution to Earnings ($/1000 units)</td>
<td>130</td>
<td>100</td>
</tr>
</tbody>
</table>

There are uncertainties that will influence his decision. 1. The total available assembly hours of next month could be 8000 or 10,000, with 50/50 chance. 2. The total available molding hours of next month could be either 21,000 or 25,000 at 50% possibility for each case.

*This is a slightly different version of Example 7.3 in the book of Bertsimas and Freund (2000).*
The Two-stage Stochastic Programming Formulation

Decision Variables: \( x \): the steel to purchase now; \( w_i, p_i \): production plan under scenario \( i = 1, 2, 3, 4 \).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Assembly Hours</th>
<th>Molding Hours</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8000</td>
<td>25000</td>
<td>.25</td>
</tr>
<tr>
<td>2</td>
<td>8000</td>
<td>21000</td>
<td>.25</td>
</tr>
<tr>
<td>3</td>
<td>10000</td>
<td>25000</td>
<td>.25</td>
</tr>
<tr>
<td>4</td>
<td>10000</td>
<td>21000</td>
<td>.25</td>
</tr>
</tbody>
</table>
We solve the problem in format (2.1) as follows.

\[
\begin{align*}
\text{min} & \quad 58x - \sum_{i=1}^{4} .25(130w_i + 100p_i) \\
\text{s.t.} & \quad w_1 + p_1 \leq 25000 \quad \text{(Mold constraint for scenario 1)} \\
& \quad .3w_1 + .5p_1 \leq 8000 \quad \text{(Assembly constraint for scenario 1)} \\
& \quad -x + 1.5w_1 + p_1 \leq 0 \quad \text{(Steel constraint for scenario 1)} \\
& \quad w_2 + p_2 \leq 21000 \quad \text{(Mold constraint for scenario 2)} \\
& \quad .3w_2 + .5p_2 \leq 8000 \quad \text{(Assembly constraint for scenario 2)} \\
& \quad -x + 1.5w_2 + p_2 \leq 0 \quad \text{(Steel constraint for scenario 2)} \\
& \quad w_3 + p_3 \leq 25000 \quad \text{(Mold constraint for scenario 3)} \\
& \quad .3w_3 + .5p_3 \leq 10000 \quad \text{(Assembly constraint for scenario 3)} \\
& \quad -x + 1.5w_3 + p_3 \leq 0 \quad \text{(Steel constraint for scenario 3)} \\
& \quad w_1 + p_1 \leq 21000 \quad \text{(Mold constraint for scenario 4)} \\
& \quad .3w_4 + .5p_4 \leq 10000 \quad \text{(Assembly constraint for scenario 4)} \\
& \quad -x + 1.5w_4 + p_4 \leq 0 \quad \text{(Steel constraint for scenario 4)} \\
& \quad x, w_i, p_i \geq 0, \ i = 1, \ldots, 4.
\end{align*}
\]
The solutions are as follows. $x = 27,250$, minimal cost $= -802,833$, and the production plans under various scenarios are as follows.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$w_i$</th>
<th>$p_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12,500</td>
<td>8,500</td>
</tr>
<tr>
<td>2</td>
<td>12,500</td>
<td>8,500</td>
</tr>
<tr>
<td>3</td>
<td>8,056</td>
<td>15,167</td>
</tr>
<tr>
<td>4</td>
<td>12,500</td>
<td>8,500</td>
</tr>
</tbody>
</table>
The SO-LDR Model

We compare the solution with our SOC reformulation with linear decision rule. Here we choose $X = \{x : x \geq 0\}$ and have

$$d = \begin{bmatrix} -130 \\ -100 \\ 0 \\ 0 \end{bmatrix}, \quad y = \begin{bmatrix} w \\ p \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}, \quad A = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ .3 & .5 & 0 & 1 & 0 \\ 1.5 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad b(\tilde{z}) = \begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \\ \tilde{z}_3 \end{bmatrix},$$

$$\ell = \begin{bmatrix} -21000 \\ -8000 \\ -1 \end{bmatrix}, \quad h = \begin{bmatrix} 25000 \\ 10000 \\ 1 \end{bmatrix},$$

where $\tau_1, \tau_2$ are slack variables, $\tilde{z}_3 \equiv 0$, and $\tilde{z}_1, \tilde{z}_2$ are random variables with

$$\mathbb{P}(\tilde{z}_1 = 21000) = \mathbb{P}(\tilde{z}_1 = 25000) = \mathbb{P}(\tilde{z}_2 = 8000) = \mathbb{P}(\tilde{z}_2 = 10000) = .5,$$

$$\mathbb{E}(\tilde{z}_1) = 23000, \mathbb{E}(\tilde{z}_2) = 9000, \mathbb{E}(\tilde{z}_3) = 0.$$
We have $E(\tilde{z}_1^2) = 533 \times 10^9, E(\tilde{z}_2^2) = 82 \times 10^9, E(\tilde{z}_3^2) = 0$, and

\[
A_0 = A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}, \quad A_i(i > 0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix},
\]

\[
b^0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad b^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad b^2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad b^3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\]
Based on the previous discussion, we solve the corresponding SOCP problem:

\[
\begin{align*}
\min_{x, v_0, v, V, u, s, t, \lambda, \nu} & \quad 58x + v_0 + 23000v_1 + 9000v_2 + 533 \times 10^9V_1 + 82 \times 10^9V_2 \\
\text{s. t.} & \quad \sum_{i=1}^{3} u_i - 130y_1^0 - 100y_2^0 + 25000\lambda_1 + 10000\lambda_2 + \lambda_3 - 21000\nu_1 - 8000\nu_2 + \nu_3 - v_0 \leq 0, \\
& \quad \left\| \frac{v_1 + 130y_1^1 + 100y_2^1 + \lambda_1 - \nu_1}{V_1 - u_1} \right\| \leq V_1 + u_1, \\
& \quad \left\| \frac{v_2 + 130y_1^2 + 100y_2^2 + \lambda_2 - \nu_2}{V_2 - u_2} \right\| \leq V_2 + u_2, \\
& \quad \left\| \frac{v_3 + 130y_1^3 + 100y_2^3 + \lambda_3 - \nu_3}{V_3 - u_3} \right\| \leq V_3 + u_3, \\
& \quad y_1^0 + y_2^0 + y_3^0 = 0, .3y_1^0 + .5y_2^0 + y_4^0 = 0, -x + 1.5y_1^0 + y_2^0 = 0, \\
& \quad y_1^1 + y_2^1 + y_3^1 = 1, .3y_1^1 + .5y_2^1 + y_4^1 = 0, 1.5y_1^1 + y_2^1 = 0, \\
& \quad y_1^2 + y_2^2 + y_3^2 = 0, .3y_1^2 + .5y_2^2 + y_4^2 = 1, 1.5y_1^2 + y_2^2 = 0, \\
& \quad y_1^3 + y_2^3 + y_3^3 = 0, .3y_1^3 + .5y_2^3 + y_4^3 = 0, 1.5y_1^3 + y_2^3 = 1 \\
& \quad y_1^1 + y_2^1 + y_3^1 = 1, 0.3y_1^1 + 0.5y_2^1 + y_4^1 = 0, 1.5y_1^1 + y_2^1 = 0, \\
& \quad y_1^2 + y_2^2 + y_3^2 = 0, 0.3y_1^2 + 0.5y_2^2 + y_4^2 = 1, 1.5y_1^2 + y_2^2 = 0, \\
& \quad y_1^3 + y_2^3 + y_3^3 = 0, 0.3y_1^3 + 0.5y_2^3 + y_4^3 = 0, 1.5y_1^3 + y_2^3 = 1, \\
& \quad -21000s_1^1 - 8000s_2^1 + s_3^1 + 25000t_1^1 + 10000t_2^1 + t_3^1 - y_1^0 \leq 0, \\
& \quad -21000s_1^2 - 8000s_2^2 + s_3^2 + 25000t_1^2 + 10000t_2^2 + t_3^2 - y_2^0 \leq 0, \\
& \quad -21000s_1^3 - 8000s_2^3 + s_3^3 + 25000t_1^3 + 10000t_2^3 + t_3^3 - y_3^0 \leq 0, \\
& \quad -21000s_1^4 - 8000s_2^4 + s_3^4 + 25000t_1^4 + 10000t_2^4 + t_3^4 - y_4^0 \leq 0,
\end{align*}
\]
\begin{align*}
s_1^1 - t_1^1 - y_1^1 & \leq 0, \quad s_2^1 - t_2^1 - y_2^2 \leq 0, \quad s_3^1 - t_3^1 - y_1^3 \leq 0, \\
s_1^2 - t_1^2 - y_1^1 & \leq 0, \quad s_2^2 - t_2^2 - y_2^2 \leq 0, \quad s_3^2 - t_2^2 - y_3^3 \leq 0, \\
s_1^3 - t_1^3 - y_1^1 & \leq 0, \quad s_2^3 - t_2^3 - y_2^2 \leq 0, \quad s_3^3 - t_3^3 - y_3^3 \leq 0, \\
s_1^4 - t_1^4 - y_1^4 & \leq 0, \quad s_2^4 - t_2^4 - y_2^4 \leq 0, \quad s_3^4 - t_3^4 - y_3^4 \leq 0, \\
x & \geq 0, \quad V, u, \lambda, \nu \geq 0, \quad s^k, t^k \geq 0, \quad k = 1, 2, 3, 4.
\end{align*}

The numerical results show that $x = 31,500$, minimal cost = $-940,770$. 
The SOCP Reformulation without LDR: The extreme points of the dual polyhedron are

\[ p^1 = (0, -44.4462, -77.7779)', \]
\[ p^2 = (-40.0014, 0, -59.9996)', \]
\[ p^3 = (0, 0, -100.0010)'. \]
\[
\begin{align*}
\min_{x, v_0, v_1, v_2, u_1, \lambda_1, \nu_1} & \quad 58x + v_0 + 23000v_1 + 9000v_2 + 533 \times 10^9 V_1 + 82 \times 10^9 V_2 \\
\text{s. t.} & \quad -77.7779x - v_0 + \sum_{j=1}^{3} u_j^1 + 25000\lambda_1^1 + 10000\lambda_2^1 + \lambda_3^1 - 21000\nu_1^1 - 8000\nu_2^1 + \nu_3^1 \leq 0, \\
& \quad -59.9996x - v_0 + \sum_{j=1}^{3} u_j^2 + 25000\lambda_1^2 + 10000\lambda_2^2 + \lambda_3^2 - 21000\nu_1^2 - 8000\nu_2^2 + \nu_3^2 \leq 0, \\
& \quad -100.001x - v_0 + \sum_{j=1}^{3} u_j^3 + 25000\lambda_1^3 + 10000\lambda_2^3 + \lambda_3^3 - 21000\nu_1^3 - 8000\nu_2^3 + \nu_3^3 \leq 0, \\
& \quad \left\| \begin{pmatrix} v_1 + \lambda_1^1 - \nu_1^1 \\ V_1 - u_1^1 \\ \end{pmatrix} \right\| \leq V_1 + u_1^1, \\
& \quad \left\| \begin{pmatrix} v_2 + \lambda_2^1 - \nu_2^1 + 44.4462 \\ V_2 - u_2^1 \\ \end{pmatrix} \right\| \leq V_2 + u_2^1, \\
& \quad \left\| \begin{pmatrix} v_3 + \lambda_3^1 - \nu_3^1 + 77.7779 \\ V_3 - u_3^1 \\ \end{pmatrix} \right\| \leq V_3 + u_3^1, \\
& \quad \left\| \begin{pmatrix} v_1 + \lambda_1^2 - \nu_1^2 + 40.0014 \\ V_1 - u_2^1 \\ \end{pmatrix} \right\| \leq V_1 + u_2^1, \\
& \quad \left\| \begin{pmatrix} v_2 + \lambda_2^2 - \nu_2^2 \\ V_2 - u_2^2 \\ \end{pmatrix} \right\| \leq V_2 + u_2^2, \\
& \quad \left\| \begin{pmatrix} v_3 + \lambda_3^2 - \nu_3^2 + 59.9996 \\ V_3 - u_3^2 \\ \end{pmatrix} \right\| \leq V_3 + u_3^2, \\
& \quad \left\| \begin{pmatrix} v_1 + \lambda_1^3 - \nu_3^3 \\ V_1 - u_1^2 \\ \end{pmatrix} \right\| \leq V_1 + u_3^3.
\end{align*}
\]
\[
\left\| \left( \frac{v_2 + \lambda^3 - \nu^3}{V_2 - u^3} \right) \right\| \leq V_2 + u^3, \\
\left\| \left( \frac{v_3 + \lambda^3 - \nu^3 + 100.001}{V_3 - u^3} \right) \right\| \leq V_3 + u^3, \\
x \geq 0, V_i, u_j^i, \lambda_j^i, \nu_j^i \geq 0, i, j = 1, 2, 3.
\]

The numerical results are as follows. \( x = 31,255, \)

minimal cost = -823,750.
Conclusion

Traditionally, a two-stage stochastic linear optimization problem is solved by using a risk-neutral approach, in which the mean of the second stage optimal value is used to make a decision. This approach can be replaced by a more realistic one by using a coherent risk measure of the second stage optimal value. This change has two major consequences. 1. It provides flexibility in dealing with the second stage problem, for example, it may accommodate risk-aversity in the decision making; and 2. It makes the second stage optimization problem take a form of worst-case optimization, therefore reduces the requirement for knowledge on distribution of the random variables involved. It often allows one to avoid the so-called curse of dimensionality and stay with computationally tractable problems.
Thank you!