Distributionally Robust Optimization with Decision-Dependent Ambiguity Set

Nilay Noyan

Sabancı University, Istanbul, Turkey

Joint work with

G. Rudolf, Koç University
M. Lejeune, George Washington University
Uncertainty in optimization

- **Stochastic programming** represents uncertain parameters by a random vector - a classical stochastic optimization:

\[
\min_{x \in \mathcal{X}} \mathbb{E}_P \left( G(x, \xi) \right)
\]

- Classical assumptions in stochastic programming:
  - The probability distribution of the random parameter vector is independent of decisions - exogenously given → relaxing it requires addressing endogenous uncertainty.
  
  - The "true" probability distribution of the random parameter vector is known → relaxing it requires addressing distributional uncertainty.
Endogenous uncertainty

- The underlying probability space may depend on the decisions:
  \[
  \min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbf{P}(\mathbf{x})} \left( G(\mathbf{x}, \xi(\mathbf{x})) \right)
  \]

- Decisions can affect the likelihood of underlying random future events.
  - **Example.** Pre-disaster planning – strengthening/retrofitting transportation links can reduce failure probabilities in case of a disaster (Peeta et al., 2010).

- Decisions can affect the possible realizations of the random parameters.
  - **Example.** Machine scheduling - stochastic processing times can be compressed by control decisions (Shabtay and Steiner, 2007).
Endogenous uncertainty

- Its use in stochastic programming remains a tough endeavor, and is far from being a well-resolved issue (Dupacova, 2006; Hellemo et al., 2018).

- Mainly two types of optimization problems (Goel and Grossmann, 2006):
  - decision-dependent information revelation
  - decision-dependent probabilities (literature is very sparse) → our focus

- Stochastic programs with decision-dependent probability measures
  - Straightforward modeling approach expresses probabilities as non-linear functions of decision variables and leads to highly non-linear models.
  - A large part of the literature focuses on a particular stochastic pre-disaster investment problem (Peeta et al., 2010; Laumanns et al., 2014; Haus et al., 2017).
  - Existing algorithmic developments are mostly specific to the problem structure.
In practice, the "true" probability distribution of uncertain model parameters/data may not be known.

- Access to limited information about the prob. distribution (e.g. samples).
- Future might not be distributed like the past.
- Solutions might be sensitive to the choice of the prob. distribution.

Distributionally robust optimization (DRO) is an appreciated approach (e.g., Goh and Sim, 2010; Wiesemann et al., 2014, Jiang and Guan, 2015).

- Considers a set of probability distributions (ambiguity set).
- Determines decisions that provide hedging against the worst-case distribution by solving a minimax type problem.
- An intermediate approach between stochastic programming and traditional robust optimization.
DRO - Choice of ambiguity set

- Moment-based versus statistical distance-based ambiguity sets
  - Exact moment-based sets typically do not contain the true distribution.
  - Conservative solutions: very different distributions can have the same lower moments and the use of higher moments can be impractical.

- Choice of statistical distance: (Bayraksan and Love, 2015; Rubner et al. 1998)
  Two of the more common ones: *Phi-divergence* versus *Earth Mover’s Distances*
  - Divergence distances do not capture the metric structure of realization space.
  - In some cases, phi-divergences limit the support of the measures in the set.
  - Our particular focus - *Wasserstein distance* with the desirable properties:
    - Consistency, tractability, etc.
A general class of Earth Mover’s Distances (EMDs)

\[ \Delta ([P, \xi_1], [Q, \xi_2]) = \inf \left\{ \int_{\Omega_1 \times \Omega_2} \delta(\xi(\omega_1), \xi(\omega_2)) \, P^*(d\omega_1, d\omega_2) : \begin{array}{c} P^* \in \mathcal{P}(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2), \\ \Pi_1(P^*) = P, \Pi_2(P^*) = Q \end{array} \right\} \]

- In a pair \([P, \xi] \in \mathcal{V}^m(\Omega, \mathcal{A}), \xi : \Omega \to \mathbb{R}^m\) is a rand. var. on the prob. space \((\Omega, \mathcal{A}, P)\)
- \(\delta\): a measure of dissimilarity (or distance) between real vectors (transportation cost)
- For any two measurable spaces \((\Omega_1, \mathcal{A}_1)\) and \((\Omega_2, \mathcal{A}_2)\), the function \(\delta\) induces an EMD
- Minimum-cost transportation plan

\[ P^* : \text{transportation plan} \]

(empirical dist.)
A general class of Earth Mover’s Distances

- Transportation problem – discrete case: \( \delta_{ij} = \delta(\xi(\omega^i), \xi(\omega^j)) \) for \( i, j \in [n] \)

\[
\min_{\gamma \in \mathbb{R}^{n \times n}_+} \left\{ \sum_{i \in [n]} \sum_{j \in [n]} \delta_{ij} \gamma_{ij} : \sum_{j \in [n]} \gamma_{ij} = p^i \quad \forall \; i \in [n], \sum_{i \in [n]} \gamma_{ij} = q^j \quad \forall \; j \in [n] \right\}
\]

- Wasserstein-\( p \) metric: \( W_p ([\mathbb{P}_1, \xi_1], [\mathbb{P}_2, \xi_2]) = \Delta^p ([\mathbb{P}_1, \xi_1], [\mathbb{P}_2, \xi_2])^{\frac{1}{p}} \), where \( \Delta^p \) is the EMD induced by \( \delta^p(x_1, x_2) = \|x_1 - x_2\|_p^p \)

- Total variation distance (also a phi-divergence distance); the EMD induced by the discrete metric

\[
\delta(x_1, x_2) = \begin{cases} 
0 & \text{if } x_1 = x_2 \\
1 & \text{if } x_1 \neq x_2.
\end{cases}
\]
DRO - Decision-dependent ambiguity set

- Incorporate distributional uncertainty into decision problems via EMD balls centered on a nominal random vector
  \[[\mathbb{P}, \xi] \in \mathcal{V}^m(\Omega, \mathcal{A}) = \mathcal{P}(\Omega, \mathcal{A}) \times \mathcal{L}^m(\Omega, \mathcal{A})\]

- Continuous EMD ball: ambiguity *both in probability measure and realizations*
  \[\mathcal{B}_{\delta, \kappa}([\mathbb{P}, \xi]) = \{\zeta \in \mathcal{L}^m([0, 1], \mathcal{A}_B) : \Delta ([\mathbb{P}, \xi], [\mathbb{B}, \zeta]) \leq \kappa\}\]

- Discrete EMD ball: the probability measure can change while the realization mapping \(\xi\) is fixed
  \[\mathcal{B}_{\delta, \kappa}^{\xi} (\mathbb{P}) = \{Q \in \mathcal{P}(\Omega, \mathcal{A}) : \Delta ([\mathbb{P}, \xi], [Q, \xi]) \leq \kappa\}\]
DRO with decision-dependent ambiguity set

Continuous EMD ball case:
\[
\min_{\mathbf{x} \in \mathcal{X}} \sup_{\zeta \in \mathcal{B}_{\delta, \kappa}([\mathbb{P}(\mathbf{x}), \xi(\mathbf{x})])} \mathbb{E}_{\mathcal{B}}(G(\mathbf{x}, \zeta))
\]

Discrete EMD ball case:
\[
\min_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbb{Q} \in \mathcal{B}_{\delta, \kappa}^{\xi(\mathbf{x})}(\mathbb{P}(\mathbf{x}))} \mathbb{E}_{\mathbb{Q}}(G(\mathbf{x}, \xi(\mathbf{x})))
\]

• DRO with Wasserstein distance has been receiving increasing attention
  ➢ See, e.g., Pflug and Wozabal, 2007; Zhao and Guan, 2015; Gao and Kleywegt, 2016; Esfahani and Kuhn, 2018; Luo and Mehrotra, 2017; Blanchet and Murthy, 2016.

• Using a decision-dependent ambiguity set: an almost untouched research area until recently

• A very recent interest on a related concept in the context of robust optimization
  ➢ Lappas and Gounaris, 2018, Nohadani and Sharma, 2018; using decision-dependent uncertainty sets.
Risk-averse variants

Continuous EMD ball case:

\[
\min_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{B}_{\delta, \kappa}([\mathbb{P}(x), \xi(x)])} \rho(G(x, \zeta))
\]

Discrete EMD ball case:

\[
\min_{x \in \mathcal{X}} \sup_{Q \in \mathcal{B}_{\delta, \kappa}^{\xi(x)}(\mathbb{P}(x))} \rho([Q, G(x, \xi(x))])
\]

- Incorporating risk is crucial for rarely occurring events such as disasters.
- Law invariant coherent risk measures defined on a standard $L_p$ space.
- Any such risk measure can be naturally extended to $p$-integrable random variables defined on an arbitrary probability space

\[
\rho([\mathbb{P}, X]) = \rho(X) = \rho\left(F_X^{(-1)}\right)
\]

- Our main focus: Conditional value-at-risk (Rockafellar and Uryasev, 2000).
Theory of risk functionals

- A risk functional $\rho$ assigns to a random variable a scalar value, providing a direct way to define stochastic preference relations: $\rho(G(x_1)) \leq \rho(G(x_2))$

- Desirable properties of risk measures, such as law invariance and coherence, have been axiomized starting with the work of Artzner et al. (1999).

- **Law invariance**: Functionals that depend only on distributions of random vars.

- **Coherence** (smaller values of risk measures are preferred):
  - Monotonicity: $X \leq Y$ a.s. $\Rightarrow \rho(X) \leq \rho(Y)$
  - Translation equivariance: $\rho(X + \lambda) = \rho(X) + \lambda$
  - Convexity: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda \rho(X) + (1-\lambda)\rho(Y)$ for $\lambda \in [0,1]$
  - Positive homogeneity: $\rho(\lambda X) = \lambda \rho(X)$ for $\lambda \geq 0$

- **CVaR** serves as a fundamental building block for other law invariant coherent risk measures (Kusuoka, 2001); supremum of convex combinations of CVaR at various confidence levels.
Value-at-risk ($\alpha$-quantile): $\text{VaR}_{0.95}(V)$ is exceeded only with a small probability of at most 0.05.

If unlucky (5% worst outcomes), the expected loss is $\text{CVaR}_{0.95}(V)$ (shaded area).

Alternative representations – Discrete case ($v_i$ with prob $p_i$, $i \in [n]$):

$$\text{CVaR}_\alpha(V) = \min \left\{ \eta + \frac{1}{1-\alpha} \mathbb{E}(V - \eta^+), \eta \in \mathbb{R} \right\}$$

$$= \max \left\{ \frac{1}{1-\alpha} \sum_{i \in [n]} v_i \beta_i : \sum_{i \in [n]} \beta_i = 1 - \alpha, 0 \leq \beta_i \leq p_i, \forall i \in [n] \right\}$$

$$= \frac{1}{1-\alpha} \int_{\alpha}^{1} \text{VaR}_\alpha(V) \, d\alpha \quad \rightarrow \text{A weighted sum of the least favorable outcomes!}$$
Formulations - Continuous EMD ball case

- Robustification of risk measures

  - Outcome mapping has a \textit{bilinear structure}: \( G(x, \xi) = \xi^T v(x), \quad v : \mathcal{X} \to \mathbb{R}^m \)
  
  - Law invariant convex risk measure \( \rho : L_p \to \mathbb{R} \) is \textit{well-behaved} with factor \( C \).
  
  - Wasserstein-\( p \) ball of radius \( \kappa \) centered on a random vector \([B, \xi]\)

  - Key result of Pflug et al. (2012): 
    \[
    \sup_{\xi \in B_{\delta p, \kappa p}([B, \xi])} \rho(\xi^T v) = \rho(\xi^T v) + C \kappa \|v\|_q
    \]

- Reformulation of the DRO problem under endogenous uncertainty:

  \[
  \min_{x \in \mathcal{X}} \rho \left( \xi^T (x) v(x) \right) + C \kappa \|v(x)\|_q
  \]
Robustifying risk measures in finite spaces

- The closed-form in the continuous case is not valid.
  
  **Example.** Let $\xi$ be a 2-dimensional random vector with possible realizations $(1, 0)^T$ and $(0, 1)^T$, and let $x = (1, 1)^T$. For any probability distribution $Q$:

  $$
  E_Q(x^T \xi) = 1 < 1 + \kappa \|x\|_q
  $$

- Using LP duality, the supremum involved in robustification of certain risk measures can be replaced with an equivalent minimization.

- The robustified CVaR value

  $$
  \sup \left\{ \text{CVaR}_\alpha ([Q, Z]) : Q \in B_{\delta,\kappa} (P) \right\} = \min \eta + \frac{1}{1-\alpha} \sum_{i=[n]} p^i v^i + \frac{1}{1-\alpha} \kappa \tau 
  $$

  subject to:

  $$
  v^i \geq z^j - \eta - \delta^{ij} \tau, \quad \forall i, j \in [n] \\
  v \in \mathbb{R}^n_+, \quad \tau \in \mathbb{R}_+.
  $$
Robustification: continuous vs. discrete balls

- A simple illustrative portfolio optimization with three equally weighted assets

- Nominal distribution:
  - Ten equally likely scenarios
  - Randomly generated losses

- Robustified $\text{CVaR}_{0.5}$ of portfolio loss
  - Ambiguity set: Wasserstein-1 ball
  - Varying radius $\kappa$

- Continuous ball
  - Loss realizations are ambiguous

- Discrete ball
  - Loss realizations are fixed
  - Only probabilities are ambiguous
Formulations - Discrete EMD ball case

- For $\rho=\text{CVaR}_\alpha$, minimax DRO problem as a conventional minimization:

\[
\min \eta + \frac{1}{1-\alpha} \sum_{\tau \in [n]} p^i(\tau) v^i + \frac{1}{1-\alpha} \kappa \tau
\]

s.t. \quad v^i \geq G(\tau, \xi^i(\tau)) - \eta - \delta^{ij} \tau, \quad \forall i, j \in [n]

\delta^{ij} = \delta(\xi^i(\tau), \xi^j(\tau)), \quad \forall i, j \in [n]

v \in \mathbb{R}_+^n, \quad \tau \in \mathbb{R}_+, \quad \tau \in \mathcal{X}.

- Analogous, although more complex, formulations can be obtained for a general class of coherent risk measures
  - the family of risk measures with finite Kusuoka representations.

- Provide an overview of various settings leading to tractable formulations.
**Tractable formulations - Discrete EMD ball**

\[
\begin{align*}
\min & \quad \eta + \frac{1}{1-\alpha} \sum_{i \in [n]} \rho_i(x)v_i + \frac{1}{1-\alpha} \kappa \tau \\
\text{s.t.} & \quad v_i \geq G_i - \eta & \forall i \in [n] \\
& \quad v_i \geq \max_{j \in [n]} G_j - \eta - \tau & \forall i \in [n] \\
& \quad v \in \mathbb{R}_+^n, \quad \tau \in \mathbb{R}_+, \quad x \in \mathcal{X}.
\end{align*}
\]

- Nominal realizations are decision-independent, and decision-dependent outcomes and scenario probabilities can be expressed via linear constraints
  - Quadratic program with linear constraints

- Both nominal realizations and outcomes are decision-independent
  - Using the discrete metric \( \delta(\xi^1, \xi^2) = \begin{cases} 0 & \text{if } \xi^1 = \xi^2 \\ 1 & \text{if } \xi^1 \neq \xi^2 \end{cases} \)
    
  - This metric allows to use total variation distance-based balls as ambiguity sets.
  - Still contains highly non-trivial instances of practical interest; pre-disaster planning (for strengthening a transportation network) and stochastic interdiction problems.
Nominal realizations are decision-dependent, and the decision-dependent outcomes and scenario probabilities can be expressed via linear constraints

Using the Wasserstein-1 metric: \[ \delta^{ij} = \| \xi^i(x) - \xi^j(x) \|_1 = \sum_{k \in [m]} \left| \xi^i_k(x) - \xi^j_k(x) \right| \]

- Mixed-binary quadratic program with quadratic constraints
- Make use of *comonotone structure* in the data to reduce the constraints of type (1)-(2), along with the corresponding binary and auxiliary variables.
Consider a transportation network where the links are subject to random failures in the event of a disaster.

- each link is either operational or non-operational
- the binary random variable: $\xi_l = 1$ (if link $l$ survives) and $\xi_l = 0$ if it fails.

Select the links to be strengthened to reduce their failure probabilities.

- No strengthening: $x_l = 0$ and $\sigma_l^0$: link survival prob.
- Strengthening (with cost $c_l$): $x_l = 1$ and $\sigma_l^1$: link survival prob.

Decision-dependent probabilities:

$$[P(x)](\{\xi_l = \xi_l^i\}) = \begin{cases} 
(1 - x_l)\sigma_l^0 + x_l\sigma_l^1 & \xi_l^i = 1 \\
(1 - x_l)(1 - \sigma_l^0) + x_l(1 - \sigma_l^1) & \xi_l^i = 0 
\end{cases}$$
Stochastic pre-disaster investment planning

- Improve post-disaster connectivity
  - Random outcome: weighted sum of shortest-path distances between a number of O-D pairs.

- Underlying risk-neutral stochastic program (Peeta et al. 2010):

\[
\min_{x \in \{0,1\}^L} \left\{ \sum_{i \in [n]} \sum_{k \in [K]} p_i(x) w_k Q_k(\xi_i) : \sum_{l \in [L]} c_l x_l \leq \hat{B} \right\}
\]

- Solve a shortest path problem for each O-D pair and scenario

- Key challenge: expressing the decision-dependent scenario probabilities

- A straightforward approach results in highly non-linear functions of decision variables (under independence assumption):

\[
p^i(x) = \prod_{\ell \in [L]: \xi^i_\ell = 1} \left[ (1 - x_\ell) \sigma^0_\ell + x_\ell \sigma^1_\ell \right] \prod_{\ell \in [L]: \xi^i_\ell = 0} \left[ (1 - x_\ell)(1 - \sigma^0_\ell) + x_\ell(1 - \sigma^1_\ell) \right].
\]
Stochastic pre-disaster investment planning

- Benefit from an efficient characterization of decision-dependent scenario probabilities via a set of linear constraints (Laumanns et al. 2014)

- Our proposed risk-neutral or CVaR-based DRO-extension:

  \[
  \min_{x \in X} \sup_{Q \in \mathcal{B}^{\xi}_{\delta, \kappa}(\mathbb{P}(x))} \rho \left( \sum_{k \in [K]} w_k Q_k(\xi) \right)
  \]

- A natural choice of ambiguity set – total variation distance-based EMD ball using the discrete metric:

  \[
  \delta(\xi^1, \xi^2) = \begin{cases} 
  0 & \text{if } \xi^1 = \xi^2 \\
  1 & \text{if } \xi^1 \neq \xi^2
  \end{cases}
  \]
Stochastic pre-disaster investment planning

- Reformulation: mixed-binary quadratic prog. with linear constraints

\[
\begin{align*}
\min & \quad \sum_{i \in [n]} \pi^i_L v^i + \kappa \tau \\
\text{s.t.} & \quad v^i \geq G^i, \quad \forall i \in [n] \\
& \quad v^i \geq \max_{j \in [n]} G^{j} - \tau, \quad \forall i \in [n] \\
\end{align*}
\]

Recursive distribution shaping
\[
\pi^i_L = p^i(\mathbf{x})
\]

\[
\begin{align*}
\pi^i_\ell & \leq \frac{\sigma^i_\ell}{\sigma^i_0} \pi^i_{\ell-1} + 1 - x_\ell, \quad \forall \ell \in [L], i \in [n] : \xi^i_\ell = 1 \\
\pi^i_\ell & \leq \frac{1 - \sigma^i_\ell}{1 - \sigma^i_0} \pi^i_{\ell-1} + 1 - x_\ell, \quad \forall \ell \in [L], i \in [n] : \xi^i_\ell = 0 \\
\pi^i_\ell & \leq \pi^i_{\ell-1} + x_\ell, \quad \forall \ell \in [L], i \in [n] \\
\sum_{i \in [n]} \pi^i_\ell & = 1, \quad \ell \in [L] \\
\sum_{\ell \in [L]} c_\ell x_\ell & \leq \hat{B}, \\
\mathbf{v} & \in \mathbb{R}^n, \quad \tau \in \mathbb{R}_+, \quad \mathbf{x} \in \{0, 1\}^L, \quad \pi \in \mathbb{R}^{L \times n}
\end{align*}
\]

- Realizations \( G^i = \sum_{k \in [K]} w_k Q_k(\xi^i), \ i \in [n] \); Baseline Probs.: \( \pi^i_0 = \prod_{\ell: \xi^i_\ell = 0} (1 - \sigma^i_\ell) \prod_{\ell: \xi^i_\ell = 1} \sigma^i_\ell \)
Robustification in finite spaces

- Robustified expectation

\[ \mathbb{E}^{\kappa}(Z) = \sup \left\{ \mathbb{E}_Q(Z) \, : \, Q \in B_{\delta,\kappa}^{\xi}(P) \right\} \]

\[ \min \sum_{i \in [n]} p^i v^i + \kappa \tau \]

s.t.
\[ v^i \geq z^i, \quad \forall i \in [n] \]
\[ v^i \geq \sup(Z) - \tau, \quad \forall i \in [n] \]
\[ \tau \geq 0. \]

- For the total variation distance

\[ \mathbb{E}^{\kappa}(Z) = \kappa \sup(Z) + (1 - \kappa) \text{CVaR}_\kappa(Z) \quad \text{(Jiang and Guan, Rahimian et al., 2018)} \]

\[ \min \kappa \sup(Z) + (1 - \kappa) \left( \eta - \frac{1}{1 - \kappa} \sum_{i \in [n]} p^i \hat{v}^i \right) \]

s.t.
\[ \hat{v}^i \geq z^i - \eta, \quad \forall i \in [n] \]
\[ \hat{v}^i \geq 0, \quad \forall i \in [n] \]
\[ \eta \leq \sup(Z). \]

- The change of variables

\[ \eta = \sup(Z) - \tau, \quad \hat{v}^i = v^i + \tau - \sup(Z) \text{ for } i \in [n]. \]
Robustification in finite spaces

- Robustified expectation $\mathbb{E}_\kappa(Z) = \sup \left\{ \mathbb{E}_Q(Z) : Q \in \mathcal{B}_{\delta,\kappa}(\mathcal{P}) \right\}$

\[
\min \sum_{i \in [n]} p^i v^i + \kappa \tau \\
\text{s.t.} \quad v^i \geq z^i, \quad \forall i \in [n] \\
\quad \quad v^i \geq \sup(Z) - \tau, \quad \forall i \in [n] \\
\quad \quad \tau \geq 0.
\]

- Optimum can be attained when $\tau = \sup(Z) - \text{VaR}_\kappa(Z)$

- $\text{VaR}_\kappa(Z) = z^j$ for at least one $j \in [n]$: $v^i = \sum_{j \in [n]} a^{ij} \beta^j$ with $a^{ij} := \max\{z^i, z^j\}$.

\[
\min \sum_{i \in [n]} \sum_{j \in [n]} p^i a^{ij} \beta^j + \kappa (\sup(Z) - \sum_{j \in [n]} z^j \beta^j) \\
\text{s.t.} \quad \sum_{j \in [n]} \beta^j = 1, \quad \beta \in \{0, 1\}^n.
\]
Stochastic pre-disaster investment planning

- Reformulation: mixed-binary quadratic prog. with linear constraints

\[ \min \sum_{i \in [n]} \pi^i_L v^i + \kappa \tau \]

s.t. \[ v^i \geq G^i, \quad \forall i \in [n] \]
\[ v^i \geq \max_{j \in [n]} G^j - \tau, \quad \forall i \in [n] \]

Distribution shaping constraints
\[ v \in \mathbb{R}^n, \quad \tau \in \mathbb{R}_+, \quad x \in X \]

- Towards an MIP formulation:

\[ \min_{x \in X} \sum_{i \in [n]} \sum_{j \in [n]} a^i_j \beta^j + \kappa \left( \max_{j \in [n]} G^j - \sum_{j \in [n]} G^j \beta^j \right) \]

s.t. Distribution shaping constraints
\[ \sum_{j \in [n]} \beta^j = 1, \quad \beta \in \{0, 1\}^n. \]

- McCormick envelopes and reformulation-linearization technique (Sherali and Adams, 1994); convex hull of (Gupte, et al. 2017)

\[ P := \{(z, \beta, \pi_L) \in \mathbb{R}_{+}^{n \times n} \times \{0, 1\}^n \times C \mid z = \pi_L \beta^\top, \quad \sum_{j \in [n]} \beta_j = 1\} \]
Stochastic pre-disaster investment planning

- Considering all the network configurations, the number of scenarios is impractically large: $2^L$.

- For computational tractability: utilize *scenario bundling* techniques.

- Laumanns et al. (2014) and Haus et al. (2017) propose very effective scenario bundling approaches.
  - For example, $2^{30}$ scenarios is replaced by 223 bundles for 5 O-D pairs.

- In the **DRO setting**, bundling raises an important issue:
  - An EMD ball around the reduced version of the original distribution is not equivalent to considering the reduced versions of the distributions in the EMD ball around the original distribution.
  - We proved that for our choice of the discrete metric these two ambiguity sets are the same.
Stochastic single-machine scheduling

- $L$ jobs with stochastic processing times;
  - machine breakdowns, inconsistency of the worker performance, changes in tool quality, variable setup times, etc.

- Find a non-preemptive job processing sequence before uncertain processing times are realized.

- Sequencing decision variables (linear ordering formulation):
  \[ \theta_{kl} = \begin{cases} 
  1, & \text{if task } k \text{ precedes task } l \\
  0, & \text{otherwise} 
\end{cases} \]

- The set $\mathcal{T}$ of feasible scheduling decisions:
  \[ \begin{align*}
  \theta_{ll} & = 1, & \forall l \in [L] \\
  \theta_{kl} + \theta_{lk} & = 1, & \forall k, l \in [L] : k < l \\
  \theta_{kl} + \theta_{lh} + \theta_{hk} & \leq 2, & \forall k, l, h \in [L] : k < l < h \\
  \theta & \in \{0, 1\}^{L \times L} 
\end{align*} \]
Controllable processing times

- Processing times are stochastic and can be affected by control decisions.

- $\xi_l(u)$: random processing time of job $l \in [L]$ given control decision $u \in U$

- A variety of schemes can be used to control processing times (e.g., Shabtay and Steiner, 2007)

**Control with discrete resources**: a set of $T$ control options for every job

- Set of feasible control decisions: $U \subset \left\{ u \in \{0, 1\}^{T \times L} : \sum_{t \in [T]} u_{tl} = 1 \ \forall l \in [L] \right\}$

- Option $t$ for job $l$ leads to a random processing time of $\hat{\xi}_{tl}$

$$\xi_l(u) = \sum_{t \in [T]} \hat{\xi}_{tl} u_{tl}, \ l \in [L]$$

$$\xi_l(u) = \hat{\xi}_l(1 - \sum_{t \in [T]} \hat{a}_{tl} u_{tl}) = \hat{\xi}_l \sum_{t \in [T]} a_{tl} u_{tl}, \ l \in [L]$$

**Comonotonicity**: $\xi^i_l(u) \geq \xi^j_l(u)$ or $\xi^i_l(u) \leq \xi^j_l(u)$ holds for all $u \in U$
Stochastic single-machine scheduling

- Random outcome of interest: **total weighted completion time**
  \[
  \sum_{l \in [L]} w_l \sum_{k \in [L]} \xi_k(u)\theta_{kl} = \sum_{k \in [L]} \sum_{l \in [L]} \xi_k(u)\theta_{kl}w_l = \xi^T(u)\Theta w
  \]

- The risk-averse version of our stochastic scheduling problem:
  \[
  \min_{(\theta, u) \in \mathcal{T} \times \mathcal{U}} h(u) + \rho \left( \xi^T(u)\Theta w \right)
  \]

- The **robustified risk-averse scheduling** problem – discrete ball
  \[
  \min_{(\theta, u) \in \mathcal{T} \times \mathcal{U}} h(u) + \sup_{Q \in \mathcal{B}^{\xi(u)}_{\delta, \kappa}(P)} \rho \left( \xi^T(u)\Theta w \right)
  \]
Stochastic single-machine scheduling

- Reformulation (mixed-integer quadratic program):

\[ \begin{align*}
\min & \quad h(u) + \eta + \frac{1}{1 - \alpha} \sum_{i \in [n]} p^i v^i + \frac{1}{1 - \alpha} \kappa \tau \\
\text{s.t.} & \quad v^i \geq \sum_{l \in [L]} \sum_{k \in [L]} \sum_{t \in [T]} w_t \hat{z}_{tkl} - \eta - \sum_{l \in [L]} v^{ij}_l \tau, \quad \forall i, j \in [n] \\
& \quad z_{tkl} \leq u_{tk}, \quad \forall t \in [T], \; k, l \in [L] \\
& \quad z_{tkl} \leq \theta_{kl}, \quad \forall t \in [T], \; k, l \in [L] \\
& \quad z_{tkl} \geq u_{tk} + \theta_{kl} - 1, \quad \forall t \in [T], \; k, l \in [L] \\
& \quad v^{ij}_l \leq \xi^i_l(u) - \xi^j_l(u) + M \lambda^{ij}_l, \quad \forall i, j \in [n], \; l \in [L] \\
& \quad v^{ij}_l \leq -\xi^i_l(u) + \xi^j_l(u) + M (1 - \lambda^{ij}_l), \quad \forall i, j \in [n], \; l \in [L] \\
& \quad \lambda \in \{0, 1\}^{n \times n \times L}, \quad \nu \in \mathbb{R}^{n \times n \times L}, \\
& \quad (\theta, u) \in \mathcal{J} \times \mathcal{U}, \quad \nu \in \mathbb{R}^n_+, \quad \tau \geq 0, \quad z \in [0, 1]^{T \times L \times L}.
\end{align*} \]

- Enhanced MIP formulations: Variable and constraint elimination, McCormick envelopes, and reformulation-linearization technique.
Computational performance

\[ \xi_l(u) = \hat{\xi}_l \sum_{t \in [T]} a_{tl} u_{tl}, \ l \in [L] \]

<table>
<thead>
<tr>
<th>( L )</th>
<th>( n )</th>
<th>\multicolumn{2}{c</th>
<th>}{Cost Minimizing}</th>
<th>\multicolumn{2}{c</th>
<th>}{Budget-Constrained}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CCM</td>
<td>CCM-RLT</td>
<td>CCM</td>
<td>CCM-RLT</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>8.4</td>
<td>7.4</td>
<td>8.9</td>
<td>7.4</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>52.9</td>
<td>46.6</td>
<td>61.6</td>
<td>23.5</td>
</tr>
<tr>
<td>150</td>
<td></td>
<td>271.3</td>
<td>137.0</td>
<td>226.2</td>
<td>175.5</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>567.2</td>
<td>539.9</td>
<td>615.3</td>
<td>359.4</td>
</tr>
<tr>
<td>250</td>
<td></td>
<td>1210.0</td>
<td>775.6</td>
<td>1626.2</td>
<td>451.8</td>
</tr>
<tr>
<td>300</td>
<td></td>
<td>1002.6</td>
<td>1023.4</td>
<td>1251.2</td>
<td>2077.5</td>
</tr>
<tr>
<td>400</td>
<td></td>
<td>1991.6</td>
<td>3155.5</td>
<td>1963.7</td>
<td>1742.2</td>
</tr>
<tr>
<td>500</td>
<td></td>
<td>5160.7</td>
<td>( ^{\dagger}6046.3[0.13] )</td>
<td>4496.1</td>
<td>( ^{\dagger}5783.11[0.45] )</td>
</tr>
<tr>
<td>15</td>
<td>50</td>
<td>113.1</td>
<td>43.0</td>
<td>172.2</td>
<td>34.5</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>603.3</td>
<td>190.0</td>
<td>627.8</td>
<td>155.3</td>
</tr>
<tr>
<td>150</td>
<td></td>
<td>2513.8</td>
<td>555.4</td>
<td>2460.8</td>
<td>376.1</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>5240.7</td>
<td>1221.9</td>
<td>5891.9</td>
<td>510.6</td>
</tr>
<tr>
<td>250</td>
<td></td>
<td>( ^{\dagger}6737.8[3.29] )</td>
<td>2659.1</td>
<td>( ^{\ddagger}[4.47] )</td>
<td>2135.5</td>
</tr>
<tr>
<td>300</td>
<td></td>
<td>( ^{\ddagger}[9.54] )</td>
<td>3581.7</td>
<td>( ^{\ddagger}[12.92] )</td>
<td>2524.2</td>
</tr>
<tr>
<td>400</td>
<td></td>
<td>–</td>
<td>( ^{\dagger}[1.34] )</td>
<td>–</td>
<td>5013.1</td>
</tr>
<tr>
<td>500</td>
<td></td>
<td>–</td>
<td>( ^{\dagger}[1.89] )</td>
<td>–</td>
<td>( ^{\ddagger}[10.91] )</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>2318.6</td>
<td>62.4</td>
<td>3357.8</td>
<td>71.6</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>( ^{\ddagger}[7.89] )</td>
<td>462.6</td>
<td>( ^{\ddagger}[18.87] )</td>
<td>523.0</td>
</tr>
<tr>
<td>150</td>
<td></td>
<td>( ^{\ddagger}[16.87] )</td>
<td>1326.7</td>
<td>( ^{\ddagger}[29.79] )</td>
<td>1044.5</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>–</td>
<td>2768.2</td>
<td>–</td>
<td>2697.5</td>
</tr>
<tr>
<td>250</td>
<td></td>
<td>–</td>
<td>5411.8</td>
<td>–</td>
<td>5035.7</td>
</tr>
<tr>
<td>300</td>
<td></td>
<td>–</td>
<td>( ^{\dagger}[1] )</td>
<td>–</td>
<td>6057.3</td>
</tr>
<tr>
<td>400</td>
<td></td>
<td>–</td>
<td>( ^{\dagger}[4.71] )</td>
<td>–</td>
<td>( ^{\dagger}[3.23] )</td>
</tr>
</tbody>
</table>

\( ^{\dagger} \): Each dagger sign indicates one instance hitting the time limit with an integer feasible solution.
Numerical Analysis

- Optimal objective function value (robustified CVaR$_\alpha$ of TWCT) for varying radius and budget ($L = 15$ jobs and $n = 100$ scenarios)
Optimal objective function values and solutions for a small illustrative example

- Solution G is only optimal for high values of $\kappa$ and low values of $\alpha$, while, conversely, solution C is only optimal for lower $\kappa$ and higher $\alpha$ values.

- Can express a range of risk-averse preferences that would not be possible to capture by either a “purely robust” or a “purely CVaR-based” approach.
Future avenues of research

- Investigate meaningful and tractable characterizations of decision-dependent nominal parameter realizations and/or scenario probabilities for practical applications.

- While *scenario bundling* is a very effective method of reducing problem sizes, most EMDs are not compatible with this approach.
  - The total variation metric is a notable exception.
  - Other class of outcome-based scenario distances, which give rise to EMDs that can be used in conjunction with bundling?

- For problems of practical interest where bundling methods are not applicable, one might instead consider *sampling methods* to reduce the number of scenarios.
  - Appropriate sampling approaches?
Robustified risk measures in finite spaces

\[
\rho^\kappa(Z) = \sup \left\{ \rho\left([\mathbb{Q}, Z]\right) : \mathbb{Q} \in \mathcal{B}_{\delta,\kappa}^\xi(\mathbb{P}) \right\} \quad \text{for } Z \in \mathcal{L}^1(\Omega, 2^\Omega).
\]

- Replacing the usual ordering with a parametric family of relations, and introducing a corresponding “penalty term”.

- **Definition.** The relation \( \preceq_\tau \): \( V \preceq_\tau Z \iff v^i \geq z^j - \delta^i j \tau \quad \forall \ i, j \in \left[ n \right] \).

- **Robustified expectation:** \( \mathbb{E}^\kappa(Z) = \inf \left\{ \mathbb{E}_\mathbb{P}(V) + \kappa \tau : \tau \geq 0, \ V \preceq_\tau Z \right\} \)

- **Robustified CVaR:**

\[
\text{CVaR}_{\alpha}^\kappa(Z) = \inf \left\{ \eta + \mathbb{E}_\mathbb{P} \left( \frac{1}{1-\alpha} S \right) + \frac{1}{1-\alpha} \kappa \tau : \eta \in \mathbb{R}, \ \tau \geq 0, \ S \preceq_\tau [Z - \eta]_+ \right\}
\]
Robustified risk measures in finite spaces

- CVaR serves as a fundamental building block for other law invariant coherent risk measures (Kusuoka, 2001)

- Robustified mixed CVaR:

\[
\rho_{\{\mu\}}(Z) = \int_0^1 \text{CVaR}_\alpha(Z) \, d\alpha = \sum_{\alpha \in \text{supp}(\mu)} \mu(\{\alpha\}) \text{CVaR}_\alpha(Z) = \mathbb{E}_\mu(\text{CVaR}_A([\mathbb{P}, Z]))
\]

\[
\rho^K_{\{\mu\}}(Z) = \inf \left\{ \mathbb{E}_\mu(H) + \mathbb{E}_\mathbb{P}(S) + \kappa \tau : \right. \\
\left. H \in \mathbb{R}^{[0,1]}, \tau \geq 0, S \succeq \tau \mathbb{E}_\mu \left( \frac{1}{1-A}[Z-H]^+ \right) \right\}.
\]

- Robustified finitely representable risk measures:

\[
\rho^*_\mathcal{M}(\mathbb{P}, Z) = \sup_{\mu \in \mathcal{M}} \rho_{\{\mu\}}([\mathbb{P}, Z])
\]

\[
\rho^*_\mathcal{M}(Z) = \inf \left\{ R \in \mathbb{R} : H \in \mathbb{R}^{[0,1]}, \tau \in \mathbb{R}_+^\mathcal{M}, \right. \\
S_{\mu} \succeq \tau_{\mu} \mathbb{E}_\mu \left( \frac{1}{1-A}[Z-H]^+ \right), \quad \forall \mu \in \mathcal{M}, \\
\left. R \geq \mathbb{E}_\mu(H) + \mathbb{E}_\mathbb{P}(S_{\mu}) + \kappa \tau_{\mu} \right\}
\]
Controllable processing times

- Processing times are stochastic and can be *affected by control decisions*.

- $\xi_l(u) \in L^1(\Omega, \mathcal{A})$: random processing time of job $l$ given decision $u \in U$
  - $U$: set of feasible control decisions
  - The mapping $\xi: U \rightarrow L^L(\Omega, \mathcal{A})$ for an arbitrary prob. space $(\Omega, \mathcal{A}, P)$

- A wide variety of schemes can be used to control processing times
  - **Linearly compressible processing times** (e.g., Shabtay and Steiner, 2007)
    \[
    \xi_l(u) = \hat{\xi}_l - a_l u_l; \quad \text{a special case } \xi_l(u) = \hat{\xi}_l (1 - u_l)
    \]
    \[
    U \subset \left\{ u \in \mathbb{R}^L : 0 \leq u_l \leq \text{ess inf} \frac{\hat{\xi}_l}{a_l} \quad \forall l \in [L] \right\}.
    \]

- **Control with discrete resources** (later)
## Computational performance

<table>
<thead>
<tr>
<th>L</th>
<th>n</th>
<th>NCM</th>
<th>PCM</th>
<th>PCM-RLT</th>
<th>PCM-RLT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CPLEX mipgap=2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>3036.7</td>
<td>151.6</td>
<td>149.5</td>
<td>138.8[0.84]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>††[13.87]</td>
<td>1361.3</td>
<td>4050.3</td>
<td>1618.8[0.93]</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>††[38.88]</td>
<td>††[0.78]</td>
<td>††[3.14]</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>–</td>
<td>††[10.10]</td>
<td>††[4.03]</td>
<td>–</td>
</tr>
<tr>
<td>15</td>
<td>50</td>
<td>††[0.59]</td>
<td>1375.8</td>
<td>398.5</td>
<td>40.9[1.23]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>††[25.25]</td>
<td>††[4.14]</td>
<td>††[0.42]</td>
<td>1403.4[1.9]</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>††[76.41]</td>
<td>††[22.66]</td>
<td>††[2.27]</td>
<td>†4538.4[2.48]</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>–</td>
<td>–</td>
<td>††[2.60]</td>
<td>–</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>††[18.9]</td>
<td>4570.5</td>
<td>1405.3</td>
<td>105.5[0.68]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>††[65.81]</td>
<td>††[23.54]</td>
<td>††[1.28]</td>
<td>2598.2[1.71]</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>–</td>
<td>††[43.5]</td>
<td>††[2.21]</td>
<td>†5140.6[2.28]</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>††[42.27]</td>
<td>††[19.04]</td>
<td>1670.8</td>
<td>161.7[1.95]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>††[79.21]</td>
<td>††[43.98]</td>
<td>††[0.52]</td>
<td>1156.6[1.11]</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>–</td>
<td>–</td>
<td>††[2.69]</td>
<td>–</td>
</tr>
<tr>
<td>30</td>
<td>50</td>
<td>††[57.81]</td>
<td>††[36.44]</td>
<td>4501.7</td>
<td>234[0.54]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>–</td>
<td>–</td>
<td>††[1.37]</td>
<td>3220.5[1.78]</td>
</tr>
</tbody>
</table>

†: Each dagger sign indicates one instance hitting the time limit with an integer feasible solution.
## Computational performance

### Impact of modeling parameters on performance of CCM-RLT

<table>
<thead>
<tr>
<th>( L )</th>
<th>( n )</th>
<th>( \alpha )</th>
<th>( \text{Cost Minimizing} )</th>
<th>( \text{Budget-Constrained} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.9</td>
<td>(1, 0.7)</td>
</tr>
<tr>
<td>15</td>
<td>50</td>
<td>37.8</td>
<td>43.0</td>
<td>31.7</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>179.1</td>
<td>190.0</td>
<td>100.4</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>526.0</td>
<td>555.4</td>
<td>367.8</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>902.2</td>
<td>1221.9</td>
<td>814.6</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>3155.5</td>
<td>2659.1</td>
<td>1909.5</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>2921.2</td>
<td>3581.7</td>
<td>2640.6</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>††[0.95]</td>
<td>††[1.35]</td>
<td>†5469.5[1.78]</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>80.5</td>
<td>62.4</td>
<td>62.1</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>367.7</td>
<td>462.6</td>
<td>337.3</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>972.1</td>
<td>1326.7</td>
<td>833.2</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>3146.0</td>
<td>2768.2</td>
<td>1742.4</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>5459.7</td>
<td>5411.8</td>
<td>3172.8</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>4820.1</td>
<td>††[1]</td>
<td>†5973.3[1.28]</td>
</tr>
</tbody>
</table>

†: Each dagger sign indicates one instance hitting the time limit with an integer feasible solution.