Bounds for Multistage Mixed-Integer Distributionally Robust Optimization

Güzin Bayraksan

Department of Integrated Systems Engineering
The Ohio State University

June 5, 2024

Joint work with **Francesca Maggioni**, **Daniel Faccini**, and **Ming Yang**
University of Bergamo, Italy and Ohio State University

Gratefully acknowledge support of **DOE** through Grants DE-AC02-06CH11347 (MACSER) and DE-SC0023361

MIP Workshop 2024

Outline

Multistage Stochastic Optimization

Many decision-making problems are stochastic and dynamic by nature. Examples:

Bond investment planning: How much bond(s) to borrow/lend every month, given that rates of return are uncertain.

Water resources allocation: How much water to allocate to different users every year, given that water supply and demand are uncertain.

Production, Logistics, Energy, Transportation, . . .

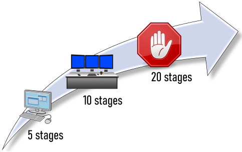
Many of these can be modeled by Multistage Stochastic Programming

Drawbacks of Multistage Stochastic Programs

- **Unknown** true distributions
  - Multistage Distributionally Robust Optimization can be used

Drawbacks of Multistage Stochastic Programs

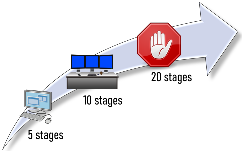
- **Unknown** true distributions
  - Multistage Distributionally Robust Optimization can be used
- **Grow exponentially** in the number of stages



- **May Lack special structure** (e.g., convexity, stagewise independence, binary state variables) that prevent efficient decomposition algorithms

Drawbacks of Multistage Stochastic Programs

- **Unknown** true distributions
  - Multistage Distributionally Robust Optimization can be used
- **Grow exponentially** in the number of stages



- **May Lack special structure** (e.g., convexity, stagewise independence, binary state variables) that prevent efficient decomposition algorithms

▶ Computationally efficient Approximations and Bounds desirable / the only method available

Different Bounding Approximations

There is a rich history of bounding approximations in stochastic programming

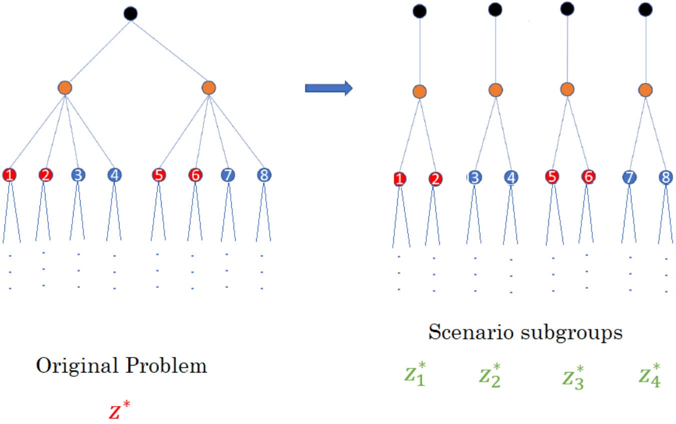
In this talk, we will use **Scenario grouping & Convolution** to generate **lower bounds** for **multistage DRO** formed with ϕ-**divergences** and **Wasserstein distance** on a finite support

Mean-CVaR (Mahmutoğullari, Çavuș & Aktürk, 2018)

Concave risk functional (Maggioni & Pflug, 2016)

Wasserstein + moment-based (Cheramin, Cheng, Jiang & Pan, 2022)

Divide & Conquer: Scenario Grouping



Outline

Multistage Distributionally Robust Optimization (DRO)

$$\min_{\mathbf{x}_0 \in \mathcal{X}_0(\boldsymbol{\xi}_0)} c_0(\mathbf{x}_0, \boldsymbol{\xi}_0) + \max_{P_1 \in \mathcal{P}_{1|\boldsymbol{\xi}^0}} \mathbb{E}_{P_1}\Bigg[\min_{\mathbf{x}_1 \in \mathcal{X}_1(\mathbf{x}_0, \boldsymbol{\xi}_1)} c_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + \max_{P_2 \in \mathcal{P}_{2|\boldsymbol{\xi}^1}} \mathbb{E}_{P_2}\Big[\dots$$

$$\dots + \max_{P_T \in \mathcal{P}_{T|\boldsymbol{\xi}^{T-1}}} \mathbb{E}_{P_T}\Big[\min_{\mathbf{x}_T \in \mathcal{X}_T(\mathbf{x}_{T-1}, \boldsymbol{\xi}_T)} c_T(\mathbf{x}_T, \boldsymbol{\xi}_T) \Big] \dots \Big] \Bigg]$$

- $\boldsymbol{\xi} := \{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T\} \in \mathbb{R}^{d_0} \times \dots \times \mathbb{R}^{d_T}$: random process
- $\boldsymbol{\xi}^t := (\boldsymbol{\xi}_0, \dots, \boldsymbol{\xi}_t)$: history of the random process
- $\mathcal{X}_t$: mixed-integer feasibility set of decision x_t
- c_t: possibly nonlinear cost functions

Multistage Distributionally Robust Optimization (DRO)

$$\min_{\mathbf{x}_0 \in \mathcal{X}_0(\boldsymbol{\xi}_0)} c_0(\mathbf{x}_0, \boldsymbol{\xi}_0) + \max_{P_1 \in \mathcal{P}_{1|\boldsymbol{\xi}^0}} \mathbb{E}_{P_1}\Bigg[\min_{\mathbf{x}_1 \in \mathcal{X}_1(\mathbf{x}_0, \boldsymbol{\xi}_1)} c_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + \max_{P_2 \in \mathcal{P}_{2|\boldsymbol{\xi}^1}} \mathbb{E}_{P_2}\Big[\dots$$
$$\dots + \max_{P_T \in \mathcal{P}_{T|\boldsymbol{\xi}^{T-1}}} \mathbb{E}_{P_T}\Big[\min_{\mathbf{x}_T \in \mathcal{X}_T(\mathbf{x}_{T-1}, \boldsymbol{\xi}_T)} c_T(\mathbf{x}_T, \boldsymbol{\xi}_T) \Big] \dots \Big] \Bigg]$$

- Assume $\boldsymbol{\xi}$ has finitely many realizations, so we can represent the stochastic process by a scenario tree

Multistage Distributionally Robust Optimization (DRO)

$$\min_{\mathbf{x}_0 \in \mathcal{X}_0(\boldsymbol{\xi}_0)} c_0(\mathbf{x}_0, \boldsymbol{\xi}_0) + \max_{P_1 \in \mathcal{P}_{1|\boldsymbol{\xi}^0}} \mathbb{E}_{P_1}\Bigg[\min_{\mathbf{x}_1 \in \mathcal{X}_1(\mathbf{x}_0, \boldsymbol{\xi}_1)} c_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + \max_{P_2 \in \mathcal{P}_{2|\boldsymbol{\xi}^1}} \mathbb{E}_{P_2}\Big[\dots$$
$$\dots + \max_{P_T \in \mathcal{P}_{T|\boldsymbol{\xi}^{T-1}}} \mathbb{E}_{P_T}\Big[\min_{\mathbf{x}_T \in \mathcal{X}_T(\mathbf{x}_{T-1}, \boldsymbol{\xi}_T)} c_T(\mathbf{x}_T, \boldsymbol{\xi}_T) \Big] \dots \Big] \Bigg]$$

- $\mathcal{P}_{t|\boldsymbol{\xi}^{t-1}}$: conditional ambiguity set of distributions

Let us first focus on Two-Stage DRO

Ambiguity Set

$\mathcal{P}_{t|\boldsymbol{\xi}^{t-1}} \rightarrow \mathcal{P}$: **ambiguity set**

Ambiguity Set

$\mathcal{P}_{t|\boldsymbol{\xi}^{t-1}} \rightarrow \mathcal{P}$: **ambiguity set**

Consider two discrete distributions $Q = \{q_{\omega_i}\}_{i=1}^{|\Omega|}$ and $P = \{p_{\omega_i}\}_{i=1}^{|\Omega|}$ with $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{|\Omega|}\}$

Ambiguity Set

$\mathcal{P}_{t|\boldsymbol{\xi}^{t-1}} \;\rightarrow\; \mathcal{P}$: **ambiguity set**

Consider two discrete distributions $Q = \{q_{\omega_i}\}_{i=1}^{|\Omega|}$ and $P = \{p_{\omega_i}\}_{i=1}^{|\Omega|}$ with $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{|\Omega|}\}$

$$\mathcal{P} := \left\{ P : \Delta(P,Q) \leq \rho, \;\; \sum_{i=1}^{|\Omega|} p_{\omega_i} = 1, \;\; p_{\omega_i} \geq 0, i = 1, \ldots, |\Omega| \right\}$$

- Δ: measure of distance/similarity between distributions
- ρ: level of robustness or **radius** (nonnegative)
- We consider Δ in $\mathcal{P}$ formulated via ϕ-divergences or Wasserstein distance

Some Commonly Used ϕ-Divergences

Consider two discrete distributions $Q = \{q_{\omega_i}\}_{i=1}^{|\Omega|}$ and $P = \{p_{\omega_i}\}_{i=1}^{|\Omega|}$ with $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{|\Omega|}\}$

$$\Delta_\phi(P, Q) = \sum_{i=1}^{|\Omega|} q_{\omega_i} \phi\left(\frac{p_{\omega_i}}{q_{\omega_i}}\right)$$

Divergence	$\phi(u)$	$\phi(u), u \geq 0$	$\Delta_\phi(P, Q)$
Variation Distance	ϕ_v	$\lvert u-1 \rvert$	$\sum \lvert p_{\omega_i} - q_{\omega_i} \rvert$
Cressie-Read Power Divergence	ϕ_{CR}^θ	$\frac{1-\theta+\theta u - u^\theta}{\theta(1-\theta)}, \theta \neq 0, 1$	$\frac{1-\sum p_{\omega_i}^\theta q_{\omega_i}^{1-\theta}}{\theta(1-\theta)}, \theta \neq 0, 1$
J-Divergence	ϕ_J	$(u-1)\log u$	$\sum (p_{\omega_i} - q_{\omega_i}) \log\left(\frac{p_{\omega_i}}{q_{\omega_i}}\right)$
χ*-Divergence of order* $a > 1$	ϕ_χ^a	$\lvert u-1 \rvert^a$	$\sum q_{\omega_i} \left\lvert 1 - \frac{p_{\omega_i}}{q_{\omega_i}} \right\rvert^a$

Some Commonly Used ϕ-Divergences

Consider two discrete distributions $Q = \{q_{\omega_i}\}_{i=1}^{|\Omega|}$ and $P = \{p_{\omega_i}\}_{i=1}^{|\Omega|}$ with $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{|\Omega|}\}$

$$\Delta_\phi(P, Q) = \sum_{i=1}^{|\Omega|} q_{\omega_i} \phi\left(\frac{p_{\omega_i}}{q_{\omega_i}}\right)$$

Divergence	$\phi(u)$	$\phi(u), u \geq 0$	$\Delta_\phi(P, Q)$
Variation Distance	ϕ_v	$\lvert u-1 \rvert$	$\sum \lvert p_{\omega_i} - q_{\omega_i} \rvert$
Cressie-Read Power Divergence	ϕ_{CR}^θ	$\frac{1-\theta+\theta u-u^\theta}{\theta(1-\theta)}, \theta \neq 0, 1$	$\frac{1-\sum p_{\omega_i}^\theta q_{\omega_i}^{1-\theta}}{\theta(1-\theta)}, \theta \neq 0, 1$
J-Divergence	ϕ_J	$(u-1)\log u$	$\sum (p_{\omega_i} - q_{\omega_i}) \log\left(\frac{p_{\omega_i}}{q_{\omega_i}}\right)$
χ*-Divergence of order* $a > 1$	ϕ_χ^a	$\lvert u-1 \rvert^a$	$\sum q_{\omega_i} \left\lvert 1 - \frac{p_{\omega_i}}{q_{\omega_i}} \right\rvert^a$

Special Cases of *Cressie-Read Power Divergence*

θ	*Corresponding Divergence*	$\phi(u)$	$\phi(u), u \geq 0$	$\Delta_\phi(P, Q)$	$\phi^\theta_{CR}(u)$
2	*Modified χ^2-Distance*	$\phi_{m\chi^2}$	$(u-1)^2$	$\sum \frac{(p_{\omega_i}-q_{\omega_i})^2}{q_{\omega_i}}$	$\frac{1}{2}(u^2-2u+1)=\frac{1}{2}(u-1)^2$
$\frac{1}{2}$	*Hellinger Distance*	ϕ_H	$(\sqrt{u}-1)^2$	$\sum(\sqrt{p_{\omega_i}}-\sqrt{q_{\omega_i}})^2$	$4(\frac{1}{2}+\frac{1}{2}u-\sqrt{u})=2(1-\sqrt{u})^2$
-1	*χ^2-Distance*	ϕ_{χ^2}	$\frac{1}{u}(u-1)^2$	$\sum \frac{(p_{\omega_i}-q_{\omega_i})^2}{p_{\omega_i}}$	$\frac{1}{2}(-2+u+\frac{1}{u})=\frac{1}{2}(\sqrt{u}-\frac{1}{\sqrt{u}})^2$
$\to 1$	*Kullback-Leibler Div.*	ϕ_{KL}	$u\log u-u+1$	$\sum p_\omega \log\left(\frac{p_\omega}{q_\omega}\right)$	$u(\log u-1)+1$
$\to 0$	*Burg Entropy*	ϕ_B	$-\log u+u-1$	$\sum q_\omega \log\left(\frac{q_\omega}{p_\omega}\right)$	$-\log u+u-1$

Kullback-Leibler Divergence and *Burg Entropy* are obtained by taking the limit of θ to 1 and 0, respectively.

Special Cases of *Cressie-Read Power Divergence*

θ	*Corresponding Divergence*	$\phi(u)$	$\phi(u), u \geq 0$	$\Delta_\phi(P, Q)$	$\phi^\theta_{CR}(u)$
2	*Modified χ^2-Distance*	$\phi_{m\chi^2}$	$(u-1)^2$	$\sum \frac{(p_{\omega_i}-q_{\omega_i})^2}{q_{\omega_i}}$	$\frac{1}{2}(u^2-2u+1)=\frac{1}{2}(u-1)^2$
$\frac{1}{2}$	*Hellinger Distance*	ϕ_H	$(\sqrt{u}-1)^2$	$\sum(\sqrt{p_{\omega_i}}-\sqrt{q_{\omega_i}})^2$	$4(\frac{1}{2}+\frac{1}{2}u-\sqrt{u})=2(1-\sqrt{u})^2$
-1	*χ^2-Distance*	ϕ_{χ^2}	$\frac{1}{u}(u-1)^2$	$\sum \frac{(p_{\omega_i}-q_{\omega_i})^2}{p_{\omega_i}}$	$\frac{1}{2}(-2+u+\frac{1}{u})=\frac{1}{2}(\sqrt{u}-\frac{1}{\sqrt{u}})^2$
$\to 1$	*Kullback-Leibler Div.*	ϕ_{KL}	$u\log u-u+1$	$\sum p_\omega \log\left(\frac{p_\omega}{q_\omega}\right)$	$u(\log u-1)+1$
$\to 0$	*Burg Entropy*	ϕ_B	$-\log u+u-1$	$\sum q_\omega \log\left(\frac{q_\omega}{p_\omega}\right)$	$-\log u+u-1$

Kullback-Leibler Divergence and *Burg Entropy* are obtained by taking the limit of θ to 1 and 0, respectively.

Special Cases of *Cressie-Read Power Divergence*

θ	*Corresponding Divergence*	$\phi(u)$	$\phi(u), u \geq 0$	$\Delta_\phi(P,Q)$	$\phi^\theta_{CR}(u)$
2	*Modified χ^2-Distance*	$\phi_{m\chi^2}$	$(u-1)^2$	$\sum \frac{(p_{\omega_i}-q_{\omega_i})^2}{q_{\omega_i}}$	$\frac{1}{2}(u^2-2u+1)=\frac{1}{2}(u-1)^2$
$\frac{1}{2}$	*Hellinger Distance*	ϕ_H	$(\sqrt{u}-1)^2$	$\sum(\sqrt{p_{\omega_i}}-\sqrt{q_{\omega_i}})^2$	$4(\frac{1}{2}+\frac{1}{2}u-\sqrt{u})=2(1-\sqrt{u})^2$
-1	*χ^2-Distance*	ϕ_{χ^2}	$\frac{1}{u}(u-1)^2$	$\sum \frac{(p_{\omega_i}-q_{\omega_i})^2}{p_{\omega_i}}$	$\frac{1}{2}(-2+u+\frac{1}{u})=\frac{1}{2}(\sqrt{u}-\frac{1}{\sqrt{u}})^2$
$\to 1$	*Kullback-Leibler Div.*	ϕ_{KL}	$u\log u-u+1$	$\sum p_\omega \log\left(\frac{p_\omega}{q_\omega}\right)$	$u(\log u-1)+1$
$\to 0$	*Burg Entropy*	ϕ_B	$-\log u+u-1$	$\sum q_\omega \log\left(\frac{q_\omega}{p_\omega}\right)$	$-\log u+u-1$

Kullback-Leibler Divergence and *Burg Entropy* are obtained by taking the limit of θ to 1 and 0, respectively.

Wasserstein Distance on a Finite Support

Consider two discrete distributions $Q = \{q_{\omega_i}\}_{i=1}^{|\Omega|}$ and $P = \{p_{\omega_j}\}_{j=1}^{|\Omega|}$ with $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{|\Omega|}\}$

$$
\begin{aligned}
\Delta_W(P, Q) = \min_{z \geq 0} & \sum_{\omega_i \in \Omega} \sum_{\omega_j \in \Omega} d_{\omega_i, \omega_j} z_{\omega_i, \omega_j} \\
\text{s.t.} & \sum_{\omega_i \in \Omega} z_{\omega_i, \omega_j} = q_{\omega_j}, \quad \forall \omega_j \in \Omega \\
& \sum_{\omega_j \in \Omega} z_{\omega_i, \omega_j} = p_{\omega_i}, \quad \forall \omega_i \in \Omega,
\end{aligned}
$$

where $d_{\omega_i, \omega_j} := ||\xi_{\omega_i} - \xi_{\omega_j}||_\varsigma$ a distance between the two scenarios ω_i and ω_j using ς-norm e.g., $\varsigma \in \{1, 2, \infty\}$.

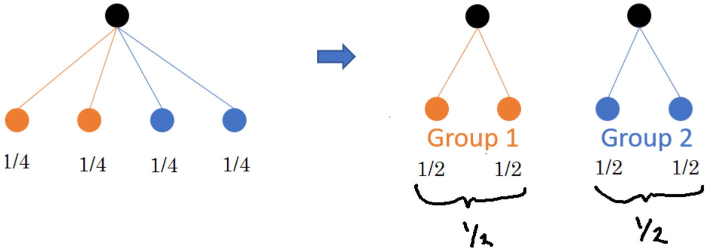
Outline

Different Bounding Approximations

In this work, we will use **Scenario grouping & convolution** to generate **lower bounds** for **multistage DRO** formed with ϕ-**divergences** and **Wasserstein distance** on a finite support

Scenario Grouping: Disjoint Partitions

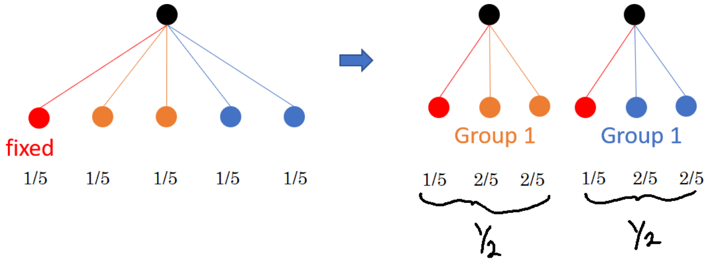
Disjoint Partitions



Can group "**similar**" or "**different**" scenarios in each group.

Scenario Grouping: Fixed Scenarios

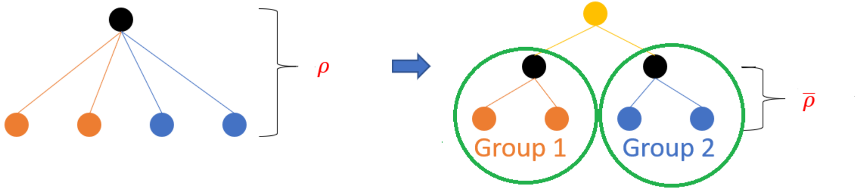
Fixed Scenarios



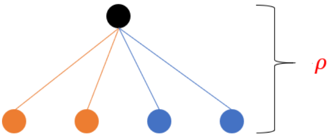
Probability of the fixed scenario changes accordingly after grouping.

Ω_f: Set of fixed scenarios

Convolution (Disjoint Partitions):
Step 1: Solve Scenario Subgroup Problems

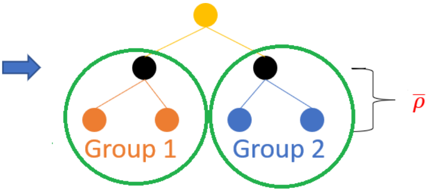


Convolution (Disjoint Partitions): Step 1: Solve Scenario Subgroup Problems



Original Problem

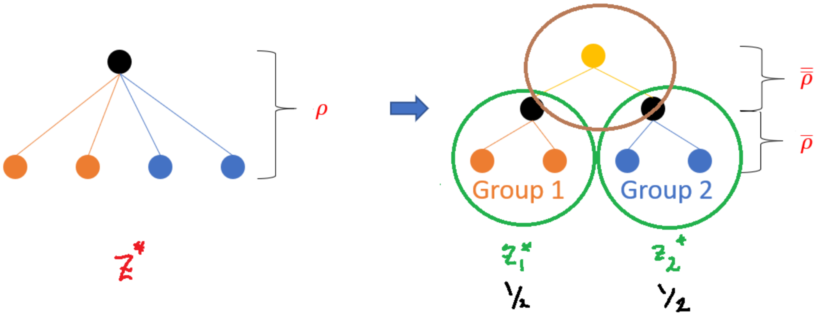
$$z^* = \min_{\mathbf{x}\in\mathcal{X}} \max_{P\in\mathcal{P}_\rho} \mathbb{E}_P\left[c(\mathbf{x},\xi)\right]$$



Step 1: Solve each scenario subgroup problem

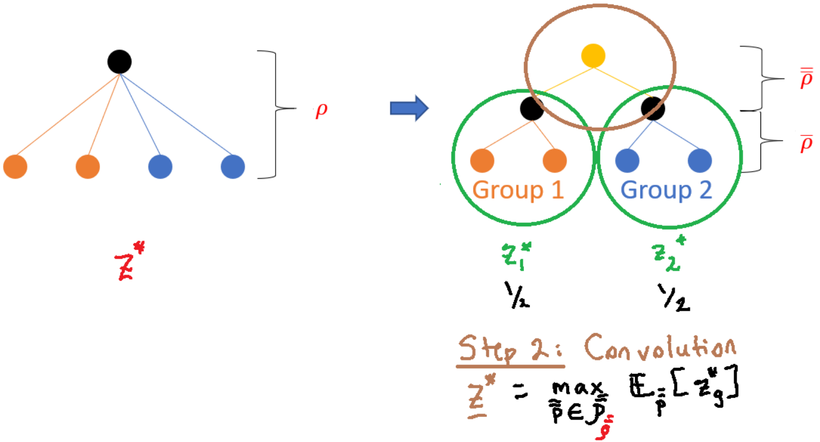
$$z_1^* = \min_{x\in X} \max_{\bar{p}\in\bar{\mathcal{P}}_{\bar{\rho}_1}} E_{\bar{p}}[c(x,\xi)] \qquad z_2^* = \min_{x\in X} \max_{\bar{p}\in\bar{\mathcal{P}}_{\bar{\rho}_2}} E_{\bar{p}}[c(x,\xi)]$$

Convolution (Disjoint Partitions):
Step 2: Combine them through convolution

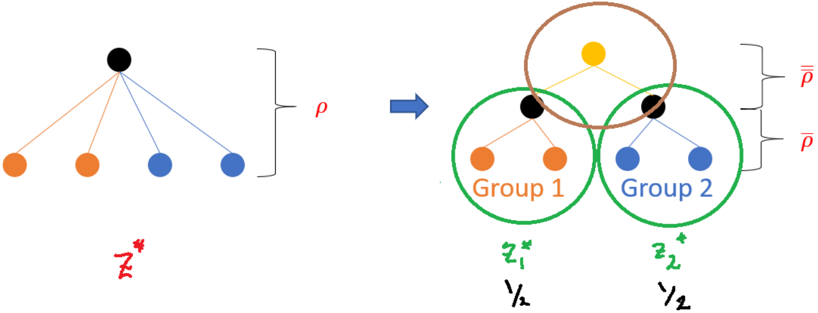


Convolution (Disjoint Partitions):

Step 2: Combine them through convolution



Convolution (Disjoint Partitions):
Step 2: Combine them through convolution

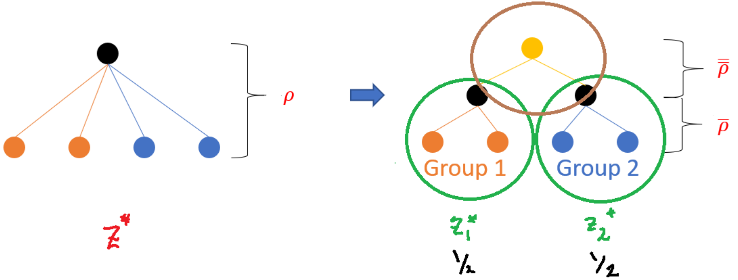


WANT:
$\underline{z}^* \leq z^*$

Step 2: Convolution

$$\underline{z}^* = \max_{\bar{\bar{p}} \in \bar{\bar{\mathcal{P}}}_{\bar{g}}} \mathbb{E}_{\bar{\bar{p}}}[z_g^*]$$

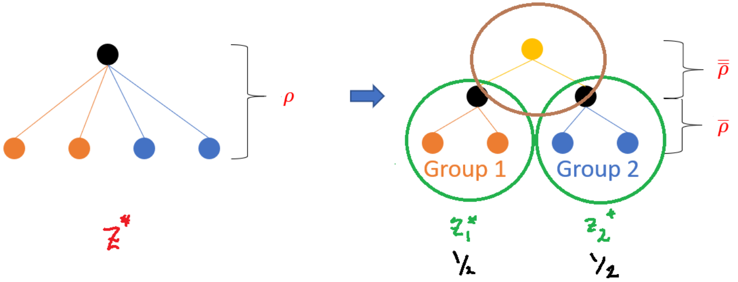
Convolution (Disjoint Partitions): Obtaining a Lower Bound



Guaranteed lower bound for **expectation** (i.e., traditional stochastic programs)

$$\mathbb{E}[z_g^*] \leq z^*$$

Convolution (Disjoint Partitions): Obtaining a Lower Bound



- **When can this method guarantee a lower bound for DRO?**
  - $\bar{\rho}_1, \bar{\rho}_2, \ldots, \bar{\rho}_G$: Radii for subgroup problem $g = 1, \ldots, G$
  - $\bar{\rho}_{max} := \max\{\bar{\rho}_1, \bar{\rho}_2, \ldots \bar{\rho}_G\}$

Lower-Bound (LB) Criteria for ϕ Divergences

Informal Theorem

1. **Cressie-Read Power Divergence:**

$$\begin{cases} \bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho & \text{when} \quad \theta \in (0,1) \\ \bar{\bar{\rho}} + \bar{\rho}_{max} - \theta(1-\theta) \cdot \bar{\bar{\rho}} \cdot \bar{\rho}_{max} \leq \rho & \text{when} \quad \theta < 0 \text{ or } \theta > 1 \text{ (and } \Omega_f = \varnothing) \\ \bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho & \text{when} \quad \theta \to 0 \text{ or } \theta \to 1, \end{cases}$$

where $\theta \neq 0,\ 1$ and Ω_f is the set of all fixed scenarios.

This result for *Cressie-Read Power Divergence* covers all special cases of its family listed in previous slide.

Lower-Bound (LB) Criteria for ϕ Divergences

Informal Theorem [cont.]

2. **Variation Distance:**

$$\bar{\bar{\rho}} \cdot \bar{\rho}_{max} + \bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho.$$

3. **J-Divergence:**

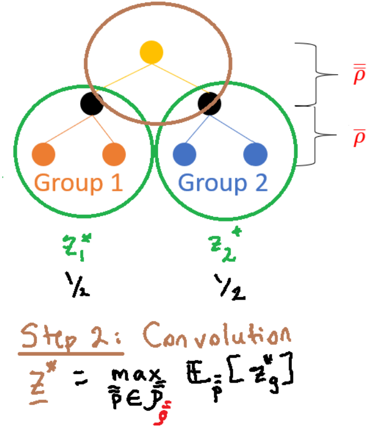
$$\bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho.$$

4. **χ-Divergence of order a:**

$$\left[(\bar{\bar{\rho}})^{\frac{1}{a}} + (\bar{\rho}_{max})^{\frac{1}{a}} + (\bar{\bar{\rho}} \cdot \bar{\rho}_{max})^{\frac{1}{a}}\right]^{a} \leq \rho \quad \text{when} \quad \Omega_f = \varnothing,$$

where $a > 1$ based on the definition and Ω_f is the set of all fixed scenarios.

Lower-Bound (LB) Criteria for Wasserstein Distance



Challenge:

How do I do Step 2 (convolution)?

What is a "distance" between groups?

Lower-Bound (LB) Criteria for Wasserstein Distance

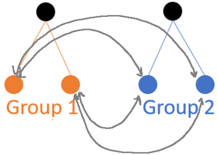
Informal Theorem [cont.]

5. **Wasserstein Distance:**
Let the distance between scenario groups be defined as

$$d_{g_1,g_2} = \begin{cases} \max\limits_{\omega_i \in \Omega^{(l)}_{g_1},\ \omega_j \in \Omega^{(l)}_{g_2}} \{d_{\omega_i,\omega_j}\} & \text{when} \quad g_1 \neq g_2,\ g_1,\ g_2 \in [m_l] \\ 0 & \text{when} \quad g_1 = g_2,\ g_1,\ g_2 \in [m_l]. \end{cases}$$

Then, the LB criteria for DRO with Wasserstein distance is as follows:

$$\bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho.$$



Outline

How to Implement the LB Criteria to Multistage Problems?

Two approaches:

1. **First-level LB** criteria for multistage problems
   - Tree can be dissected in any way
   - Convolution done "only" at the first stage
   - Applicable to ϕ-divergences
   - [Applicable to Wasserstein only if dissected at 1st stage]

2. **Multi-level LB** criteria for multistage problems
   - Tree must be dissected in a specific way
   - Convolution performed in a backward "nested" manner
   - Applicable to both Wasserstein and ϕ-divergences

First-Level LB

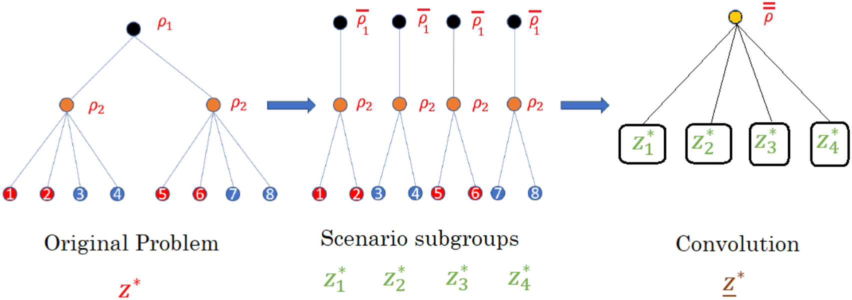
First-Level LB Multistage Stochastic Programs

After dissecting the scenario tree,

1. **Only change first-stage** ρ_1 **to** $\bar{\rho}_1$
2. For all other stages, $t = 2, \ldots, T$, keep the radii the **same** as the original problem: $\rho_2, \ldots, \rho_T$
3. Then, **combine the optimal values** of scenario groups via $\bar{\bar{\rho}}$

To obtain LB, $\bar{\rho}_1$, $\bar{\bar{\rho}}$ must satisfy above criteria

First-Level LB Multistage Stochastic Programs



Total Variation Distance: if $\bar{\bar{\rho}} \cdot \bar{\rho}_1 + \bar{\bar{\rho}} + \bar{\rho}_1 \leq \rho_1$ then $\underline{z}^* \leq z^*$

Applicable to ϕ-**divergences**

(or Wasserstein distance with tree dissected *only* at the root node)

First-Level LB

Informal Theorem (First-Level LB)

With scenario tree dissected in any form, apply the above procedure to only the first-stage for Multistage DRO formed via ϕ-divergences (or Wasserstein distance with tree dissected *only* at the root node). If the parameters $\bar{\bar{\rho}}$ and $\bar{\rho}_{max}$ satisfy the above criteria with respect to original problem's ρ_1, then,

First-Level Scenario Grouping & Convolution provides a valid LB.

Multi-Level LB

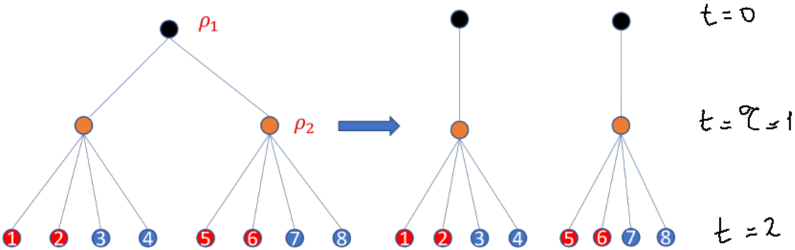
Multi-Level LB Multistage Stochastic Programs

Unfortunately, we do not have a distance between scenario groups (in the general multistage case) for Wasserstein distance.

Multi-Level LB Multistage Stochastic Programs

Unfortunately, we do not have a distance between scenario groups (in the general multistage case) for Wasserstein distance.

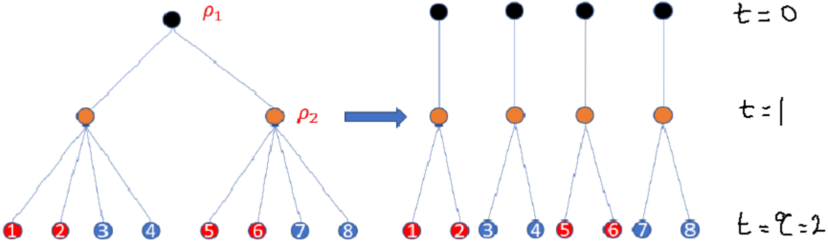
1. Let τ be the last stage the scenario tree is dissected



Multi-Level LB Multistage Stochastic Programs

Unfortunately, we do have a distance between scenario groups (in the general multistage case) for Wasserstein distance.

1. Let τ be the last stage the scenario tree is dissected



Scenario Tree Dissected Up To Stage τ

Informal Definition (Scenario Tree Dissected Up To Stage τ)

1. **Same Ancestor:** All nodes at stage τ in a given subgroup must have the same ancestor,
2. **All Children:** All children of stage τ in a given subgroup must be in that subgroup,
3. **Disjoint subgroups:** The nodes at stage τ of different subgroups must be disjoint.

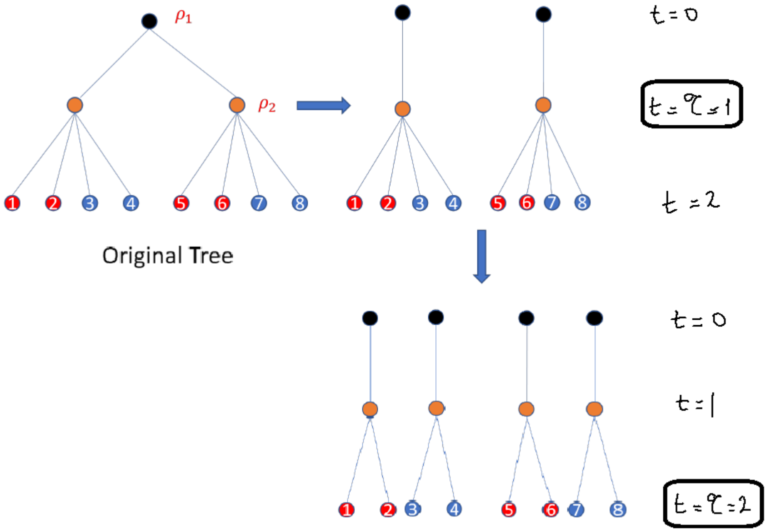
Scenario Tree Dissected Up To Stage τ

Informal Definition (Scenario Tree Dissected Up To Stage τ)

1. **Same Ancestor:** All nodes at stage τ in a given subgroup must have the same ancestor,
2. **All Children:** All children of stage τ in a given subgroup must be in that subgroup,
3. **Disjoint subgroups:** The nodes at stage τ of different subgroups must be disjoint.

- Tree dissected up to stage $\tau \in \mathcal{T} \setminus \{0, 1\}$ forms a **refinement** of a tree dissected up to stage $\tau - 1$

A Nested Refinement

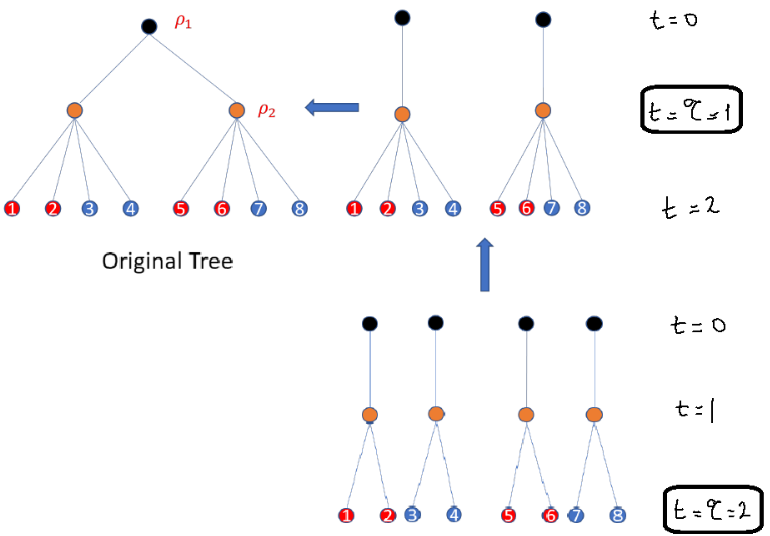


Multi-Level LB Multistage Stochastic Programs

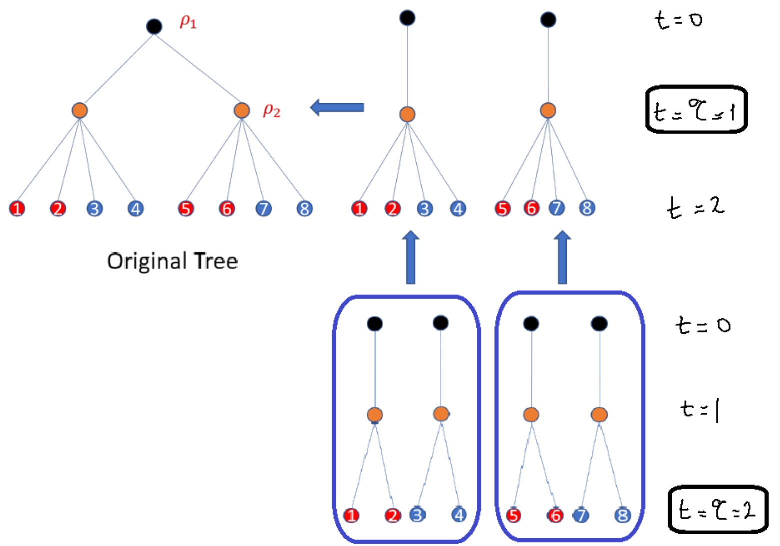
2. **Up to Stage τ:** For stages $t = 1, \ldots, \tau$, **choose radii** $\bar{\rho}_1, \ldots, \bar{\rho}_\tau$
3. **From $\tau + 1$ to End:** For $t = \tau + 1, \ldots T$, keep the **same radii** as the original problem: $\rho_{\tau+1}, \ldots, \rho_T$
4. **Combine/Convolute in a recursive way:** starting from τ up to stage $t = 1$ using $\bar{\bar{\rho}}_\tau, \ldots, \bar{\bar{\rho}}_1$
5. At each stage up to stage τ, the corresponding LB criteria must be satisfied. **Ex:** For Wasserstein distance,

$$\bar{\bar{\rho}}_t + \bar{\rho}_{t,max} \leq \rho_t \quad \text{for all } t = \tau, \ldots, 1$$

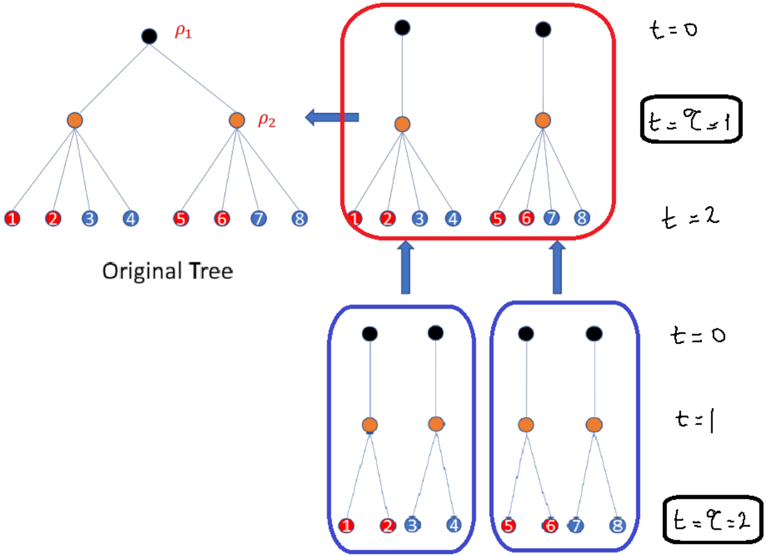
A Nested Refinement



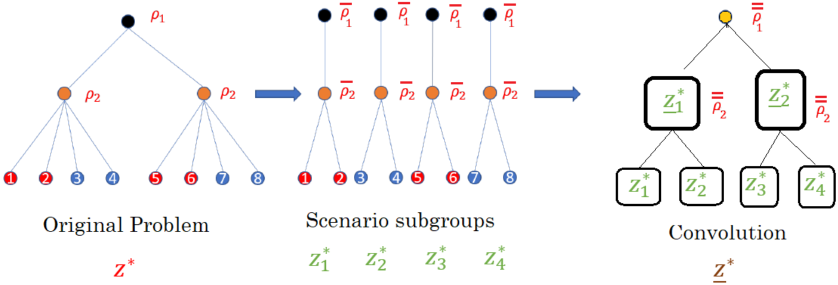
A Nested Refinement



A Nested Refinement



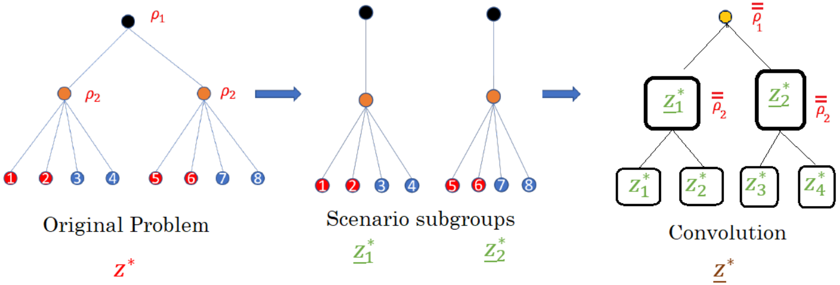
Multi-Level LB Multistage Stochastic Programs



Wasserstein Distance: if $\bar{\bar{\rho}}_t + \bar{\rho}_t \leq \rho_t$ for all $t = 1, \dots, \tau$ then $\underline{z}^* \leq z^*$

Applicable to both ϕ-**divergences** and **Wasserstein distance**

Multi-Level LB Multistage Stochastic Programs



Wasserstein Distance: if $\bar{\bar{\rho}}_t + \bar{\rho}_t \leq \rho_t$ for all $t = 1, \ldots, \tau$ then $\underline{z}^* \leq z^*$

Applicable to both ϕ-**divergences** and **Wasserstein distance**

Multi-Level LB

Informal Theorem (Multi-Level LB)

With scenario tree dissected up to stage τ, apply the above procedure to stages $t = \tau, \dots, 1$ for Multistage DRO formed via ϕ-divergences or Wasserstein distance. If the parameters $\bar{\bar{\rho}}_t$ and $\bar{\rho}_{t,max}$ satisfy the above criteria with respect to original problem's ρ_t, for all $t = 1, \dots, \tau$ then,

Multi-Level Scenario Grouping & Convolution provides a valid LB.

Outline

Multistage Mixed-Integer Production/Inventory Problem

We applied the LB methods to a multistage mixed-integer production inventory problem, modified from (Maggioni & Pflug, 2016)

The problem has

- 6 stages $t = 0, 1, \ldots T = 5$
- 540 total scenarios $(= 5 \times 4 \times 3 \times 3 \times 3)$
- Scenarios obtained by discretized time-inhomogenous exponential autoregressive with lag 1 process
- $\rho_t = 0.5$ for all $t = 1, \ldots, T$ and all ambiguity sets
- Also created a two-stage variant with 100 scenarios with $\rho = 0.1$ for Variation Distance and Modified χ^2 and $\rho = 1.5$ for Wasserstein

Two-Stage Results:

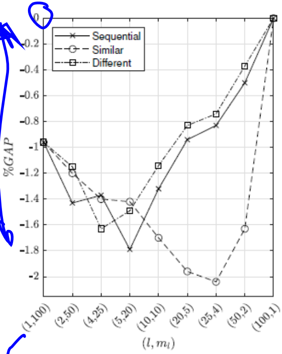
Effect of Grouping + **Best** $(\bar{\bar{\rho}}, \bar{\rho}_{max})$ **combination**

Two-Stage Results: Effect of Grouping

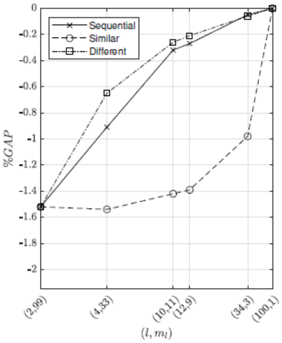
In addition to "disjoint" and "fixed" scenarios, we examine

- **Similar:** Groups get similar scenarios (e.g., all low-demand)
- **Different:** Groups have different scenarios (e.g., some low- some high-demand scenarios)
- **Sequential:** The way scenario tree was generated (e.g., random)

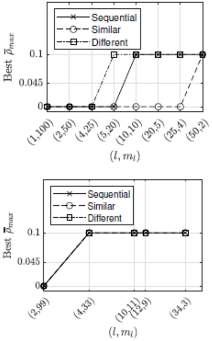
Two-Stage Results – Variation Distance



(a) Best LB – Disjoint



(b) Best LB – Fixed



(c) Best $\bar{\rho}_{max}$

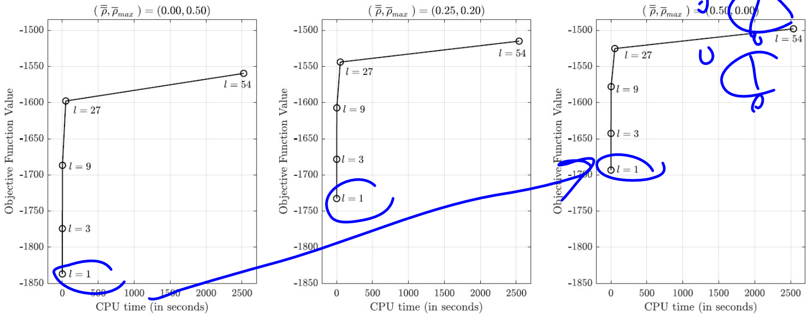
Multi-Stage Results:

First-Level LB vs. Multi-Level LB

Results on Modified χ^2 Distance: First-Level LB

We choose the combinations $(\bar{\bar{\rho}}, \bar{\rho}_{max})$ with:

$$\bar{\bar{\rho}} \in \{0.00, 0.25, 0.50\} \text{ and } \bar{\rho}_{max} = \frac{\rho_1 - \bar{\bar{\rho}}}{1+\bar{\bar{\rho}}}.$$

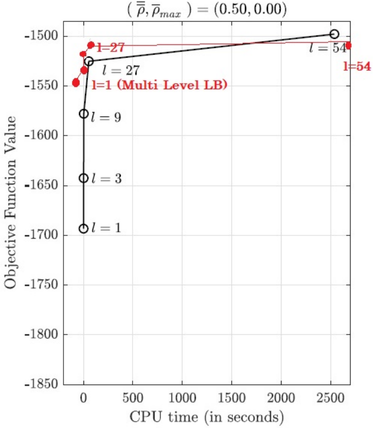


We choose subsets to be **disjoint** and follow the structure of the scenario tree with size of subgroups $l = 1, 3, 9, 27, 54$, applying first-level LB.

Best upper bound – Best LB ≈ 3%

Results on Modified χ^2 Distance: First-Level vs. Multi-Level LB

We look at only $\bar{\bar{\rho}} = 0.5, \ \bar{\rho}_{max} = 0.0$



Outline

Takeaways - I

- The **greater the number of scenarios** per subproblem, the **sharper** the obtained lower bounds
- **Fixing** certain high-cost scenarios can be **beneficial** for some DRO (e.g., Variation distance; Wasserstein distance)
- When the **size of the subgroups is small**, numerical results indicate that **multi-level** LB can be more effective than first-level LB

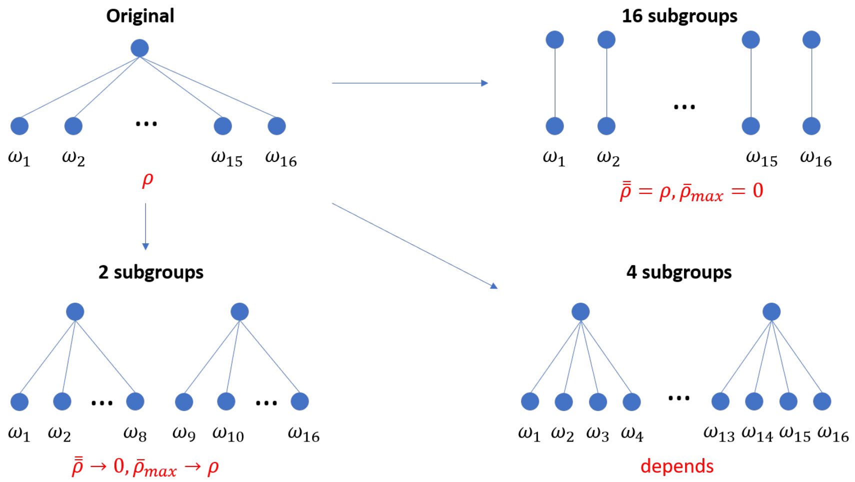
Takeaways - II

- When the **dispersion within subgroups increases**,
  - Different (or Sequential) strategies
  - Fixing worst-case scenario
  - Size of subgroups at the point where bounding is applied is larger

  more importance should be assigned to $\bar{\rho}_{max}$ (i.e., $\boldsymbol{\bar{\rho}_{max} \nearrow \rho}$) at the expense of $\bar{\bar{\rho}}$ (i.e., $\boldsymbol{\bar{\bar{\rho}} \searrow 0}$)

- and vice versa

Takeaways - III (disjoint scenarios)



Can these bounds be used in an algorithmic way?

Can these bounds be used in an algorithmic way?

Yes.

Ahmed (2013)
Mahmutoğullari, Çavuş & Aktürk (2019)
Deng et al. (2021)

Can these bounds be used in an algorithmic way?

Yes.

Ahmed (2013)
Mahmutoğullari, Çavuş & Aktürk (2019)
Deng et al. (2021)

Assume first-stage variables are binary

$\mathcal{B}$: set of first-stage variables

Solution Algorithm

0 Initialize $UB = +\infty$, $LB = -\infty$; $\mathcal{C} = \emptyset$; A Dissection of Scenario Tree

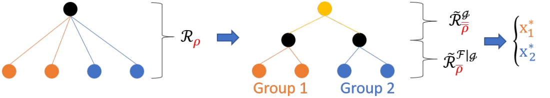
While $UB > LB$ and $\mathcal{B} \setminus \mathcal{C} \neq \emptyset$

Solution Algorithm

0 Initialize $UB = +\infty$, $LB = -\infty$; $\mathcal{C} = \emptyset$; A Dissection of Scenario Tree

While $UB > LB$ and $\mathcal{B} \setminus \mathcal{C} \neq \emptyset$

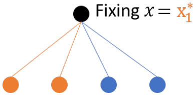
1 Scenario Decomposition (lower bound) on set $\mathcal{B} \setminus \mathcal{C}$:

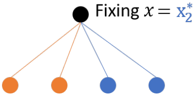


Add newly obtained solutions to $\mathcal{C}$;

Update LB

2 Evaluation (upper bound):





Update UB

Solution Algorithm: No-Good Cuts

Each group subproblem (MINLP) is solved over updated set

$$\mathcal{B} \backslash \mathcal{C} = \left\{ x \in \mathcal{B} : \sum_{i : x_i' = 1} (1 - x_i) + \sum_{i : x_i' = 0} x_i \geq 1 \; \forall x' \in \mathcal{C} \right\},$$

Food for Thought

- Can the algorithm be improved (e.g., specialized for some problems)?
- Is there a way to optimally partition the scenario tree?

  Ryan, Ahmed, Dey, Rajan, Musselman, Watson, (2020) "Optimization-Driven Scenario Grouping"

Food for Thought

- Can the algorithm be improved (e.g., specialized for some problems)?
- Is there a way to optimally partition the scenario tree?

  Ryan, Ahmed, Dey, Rajan, Musselman, Watson, (2020) "Optimization-Driven Scenario Grouping"
- Can we use **"effective" scenarios** to strengthen the bounds?
- Can we use within sampling-based methods like SDDP?

Thank you!

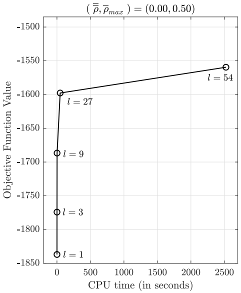
(bayraksan.1@osu.edu)

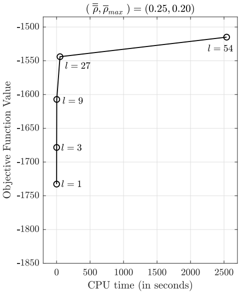
Bayraksan, G., Maggionni, F., Faccini, D. and M. Yang,
"Bounds for Multistage Mixed-Integer Distributionally Robust Optimization,"
SIAM Journal on Optimization, 34(1): 682–717, 2024

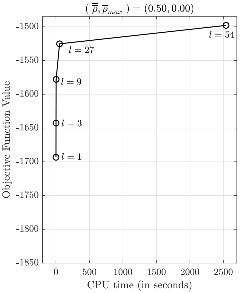
Results on Modified χ^2 Distance

We choose the combinations $(\bar{\bar{\rho}}, \bar{\rho}_{max})$ with:

$$\bar{\bar{\rho}} \in \{0.00, 0.25, 0.50\} \text{ and } \bar{\rho}_{max} = \frac{\rho_1 - \bar{\bar{\rho}}}{1+\bar{\bar{\rho}}}.$$



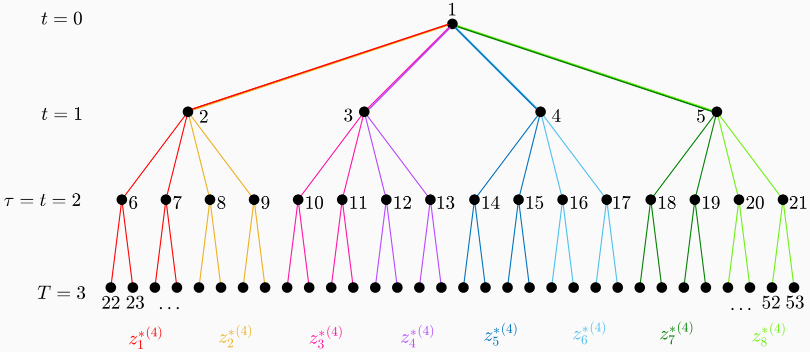




We choose subsets to be **disjoint** and follow the structure of the scenario tree with size of subgroups $l = 1, 3, 9, 27, 54$.

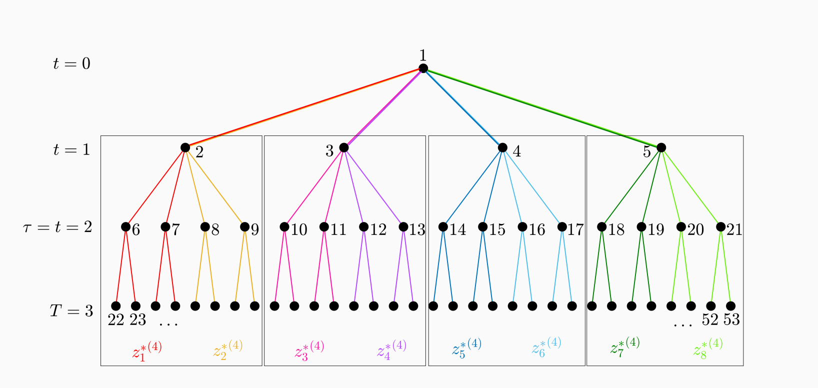
Best upper bound – Best LB ≈ 3%

LB Criteria for Multistage DRO with Wasserstein Distance



$$\zeta_{LB} := \left[z_1^{*(4)}, z_2^{*(4)}, z_3^{*(4)}, z_4^{*(4)}, z_5^{*(4)}, z_6^{*(4)}, z_7^{*(4)}, z_8^{*(4)} \right]$$

LB Criteria for Multistage DRO with Wasserstein Distance

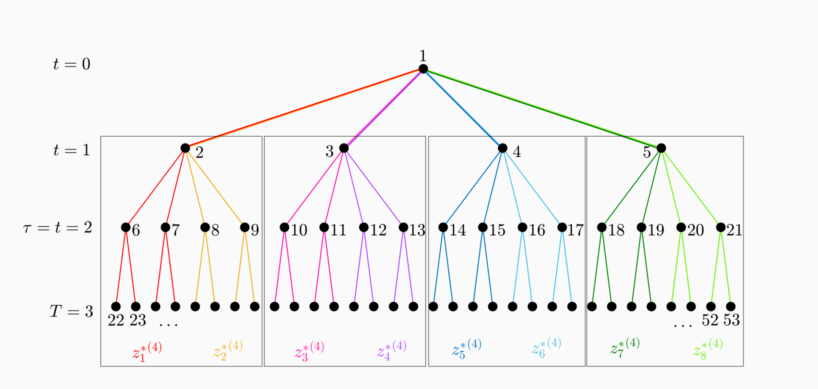


Distance Between Scenario Groups

$$d_{g_1,g_2} := \begin{cases} \max_{i \in \Omega^{(l)}_{\tau,g_1},\, j \in \Omega^{(l)}_{\tau,g_2}} \{d_{i,j}\} & \text{when} \quad g_1 \neq g_2, \\ 0 & \text{when} \quad g_1 = g_2, \end{cases}$$

where g_1 and g_2 are chosen s.t. $\forall n_1 \in \Omega^{(l)}_{\tau,g_1}, n_2 \in \Omega^{(l)}_{\tau,g_2} : a(n_1) = a(n_2)$

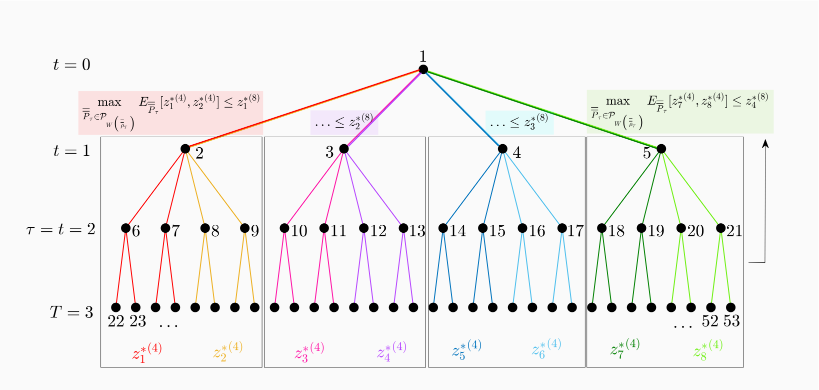
LB Criteria for Multistage DRO with Wasserstein Distance



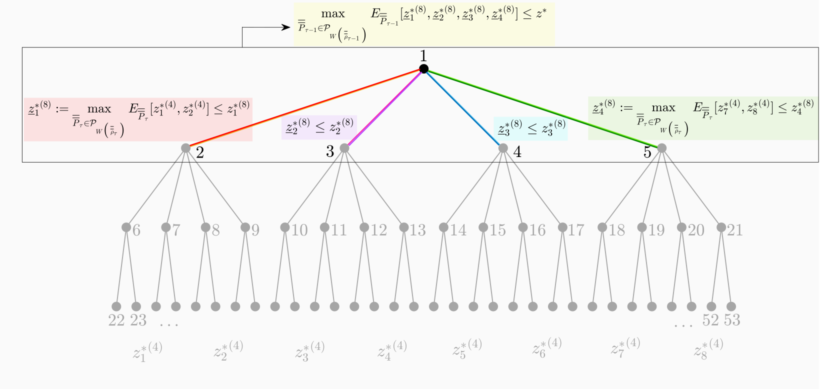
Example: Distance Between Groups 1 and 2

$$d_{1,2} := \max\left\{||\xi_6 - \xi_8||_\varsigma, ||\xi_6 - \xi_9||_\varsigma, ||\xi_7 - \xi_8||_\varsigma, ||\xi_7 - \xi_9||_\varsigma\right\}.$$

LB Criteria for Multistage DRO with Wasserstein Distance



LB Criteria for Multistage DRO with Wasserstein Distance



Distances Between Groups

$$d_{1,2} := ||\xi_2 - \xi_3||_\varsigma, \quad d_{1,3} := ||\xi_2 - \xi_4||_\varsigma,$$
$$d_{1,4} := ||\xi_2 - \xi_5||_\varsigma, \quad d_{2,3} := ||\xi_3 - \xi_4||_\varsigma,$$
$$d_{2,4} := ||\xi_3 - \xi_5||_\varsigma, \quad d_{3,4} := ||\xi_4 - \xi_5||_\varsigma.$$