Effective Scenarios in Multistage Distributionally Robust Optimization with Total Variation Distance

Güzin Bayraksan

Department of Integrated Systems Engineering
The Ohio State University

April 4, 2019

Joint work with Hamed Rahimian (Northwestern University) and
Tito Homem-de-Mello (Universidad Adolfo Ibañez)

East Coast Optimization Meeting 2019 @ Dept. of Mathematical Sciences
George Mason University

Outline

Outline

Stochastic Dynamic Programs

Many decision-making problems are stochastic and dynamic by nature. For example,

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

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

Dynamics

$$x_1 \rightsquigarrow \xi_2 \rightsquigarrow x_2$$

Dynamics

$$x_1 \rightsquigarrow \xi_2 \rightsquigarrow x_2 \rightsquigarrow \xi_3 \rightsquigarrow x_3 \rightsquigarrow \ldots \rightsquigarrow \xi_T \rightsquigarrow x_T$$

Dynamics

$$x_1 \rightsquigarrow \xi_2 \rightsquigarrow x_2 \rightsquigarrow \xi_3 \rightsquigarrow x_3 \rightsquigarrow \ldots \rightsquigarrow \xi_T \rightsquigarrow x_T$$

- Stochastic programming, stochastic optimal control, Markov decision processes are ways to model these problems, among others.
- We focus on a particular class of problems:

  Multistage stochastic program (MSP)

General Formulation of MSP

$$\min_{x_1, x_2, \ldots, x_T} \mathbb{E}\left[g_1(x_1, \xi_1) + g_2(x_2, \xi_2) + \ldots + g_T(x_T, \xi_T)\right]$$
$$\text{s.t.} \quad x_t \in \mathcal{X}_t := \mathcal{X}_t(x_{[t-1]}, \xi_{[t]}), \; t = 1, 2, \ldots T,$$

where

- $\xi_{[t]}$ and $x_{[t]}$: history of stochastic process and decisions up to stage t
- $x_t := x_t(\xi_{[t]})$: decision made at each stage
- $\mathcal{X}_t := \mathcal{X}_t(x_{[t-1]}, \xi_{[t]})$: feasibility set in stage t
- $g_t(x_t, \xi_t)$: cost of decision x_t given the realized uncertainty ξ_t at stage t

General Formulation of MSP

$$\min_{x_1,x_2,\ldots,x_T} \mathbb{E}\left[g_1(x_1,\xi_1)+g_2(x_2,\xi_2)+\ldots+g_T(x_T,\xi_T)\right]$$
$$\text{s.t.} \quad x_t \in \mathcal{X}_t := \mathcal{X}_t(x_{[t-1]},\xi_{[t]}),\ t=1,2,\ldots T,$$

where

- $\xi_{[t]}$ and $x_{[t]}$: history of stochastic process and decisions up to stage t
- $x_t := x_t(\xi_{[t]})$: decision made at each stage
- $\mathcal{X}_t := \mathcal{X}_t(x_{[t-1]},\xi_{[t]})$: feasibility set in stage t
- $g_t(x_t,\xi_t)$: cost of decision x_t given the realized uncertainty ξ_t at stage t

General Formulation of MSP

$$\min_{x_1, x_2, \ldots, x_T} \mathbb{E}\left[g_1(x_1, \xi_1) + g_2(x_2, \xi_2) + \ldots + g_T(x_T, \xi_T)\right]$$
$$\text{s.t.} \quad x_t \in \mathcal{X}_t := \mathcal{X}_t(x_{[t-1]}, \xi_{[t]}), \; t = 1, 2, \ldots T,$$

where

- $\xi_{[t]}$ and $x_{[t]}$: history of stochastic process and decisions up to stage t
- $x_t := x_t(\xi_{[t]})$: decision made at each stage
- $\mathcal{X}_t := \mathcal{X}_t(x_{[t-1]}, \xi_{[t]})$: convex feasibility set in stage t
- $g_t(x_t, \xi_t)$: convex cost of decision x_t given the realized uncertainty ξ_t at stage t

General Formulation of MSP

$$\begin{aligned} \min_{x_1, x_2, \ldots, x_T} \ & \mathbb{E}\left[g_1(x_1, \xi_1) + g_2(x_2, \xi_2) + \ldots + g_T(x_T, \xi_T)\right] \\ \text{s.t.} \ & x_t \in \mathcal{X}_t := \mathcal{X}_t(x_{[t-1]}, \xi_{[t]}), \ t = 1, 2, \ldots T, \end{aligned}$$

Nested Formulation of MSP

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \mathbb{E}_{\mathbf{q}_2|\xi_{[1]}} \left[\min_{x_2 \in \mathcal{X}_2} g_2(x_2, \xi_2) + \mathbb{E}_{\mathbf{q}_3|\xi_{[2]}} \left[\dots + \mathbb{E}_{\mathbf{q}_T|\xi_{[T-1]}} \left[\min_{x_T \in \mathcal{X}_T} g_T(x_T, \xi_T) \right] \dots \right] \right]$$

- $\mathbf{q}_{t|\xi_{[t-1]}}$: conditional distribution of stage t, conditioned on $\xi_{[t-1]}$
- $\mathbb{E}_{\mathbf{q}_t|\xi_{[t-1]}}[\cdot]$: conditional expectation w.r.t. $\mathbf{q}_{t|\xi_{[t-1]}}$

Nested Formulation of MSP

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \mathbb{E}_{\mathbf{q}_2 | \xi_{[1]}} \left[\min_{x_2 \in \mathcal{X}_2} g_2(x_2, \xi_2) + \mathbb{E}_{\mathbf{q}_3 | \xi_{[2]}} \left[\ldots + \mathbb{E}_{\mathbf{q}_T | \xi_{[T-1]}} \left[\min_{x_T \in \mathcal{X}_T} g_T(x_T, \xi_T) \right] \ldots \right] \right]$$

Features/Assumptions

- Expectation is w.r.t. known joint probability distribution of $\{\xi_t\}_{t=1}^T$
- Assume ξ_t has finitely many possible realizations, so we can represent the process using a scenario tree
- Optimization is done over policies $x := [x_1, \ldots, x_T]$

Drawbacks of the Previous Model

The decision maker

1. is risk-neutral,
2. have complete information about the underlying uncertainty via a known probability distribution.

Drawbacks of the Previous Model

The decision maker

1. is risk-neutral,
2. have complete information about the underlying uncertainty via a known probability distribution.

$\longrightarrow$ What if this is not the case?

Drawbacks of the Previous Model

The decision maker

1. is risk-neutral,
2. have complete information about the underlying uncertainty via a known probability distribution.

$\longrightarrow$ What if this is not the case?

The distributionally robust version of the problem (multistage DRSP) addresses the situation where the decision maker

1. might be risk-averse,
2. might have partial information about the underlying probability distribution, e.g., from historical data and/or expert opinions.

Motivation

Address the following fundamental research questions in the context of multistage DRSP (and many other decision-making problems under uncertainty):

Q1: How do we formulate this problem?

Q2: What uncertain scenarios are *important* to a multistage DRSP model?

- How to define *important* scenarios?
- How to identify *important* scenarios?

Motivation

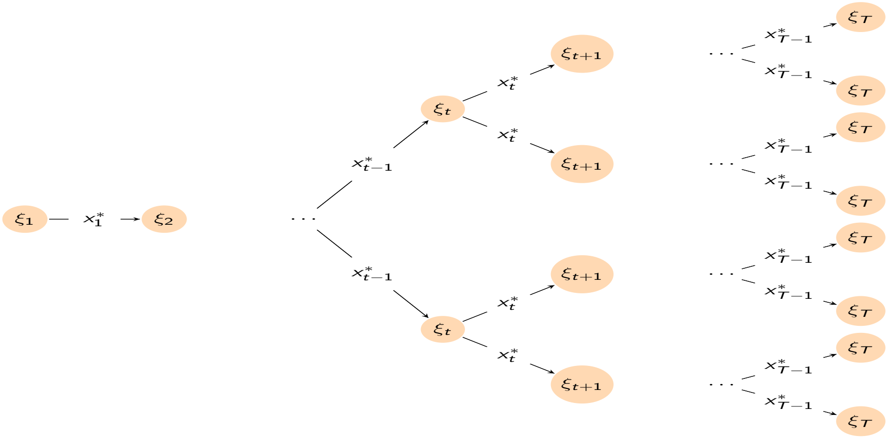
Q3: What can be inferred from *important* scenarios in real-world applications?

- Encourage decision makers to collect more accurate information surrounding these scenarios
- Help decision maker to choose an appropriate size for the ambiguity sets
- Accelerate Decomposition Algorithms
- Scenario Reduction

Outline

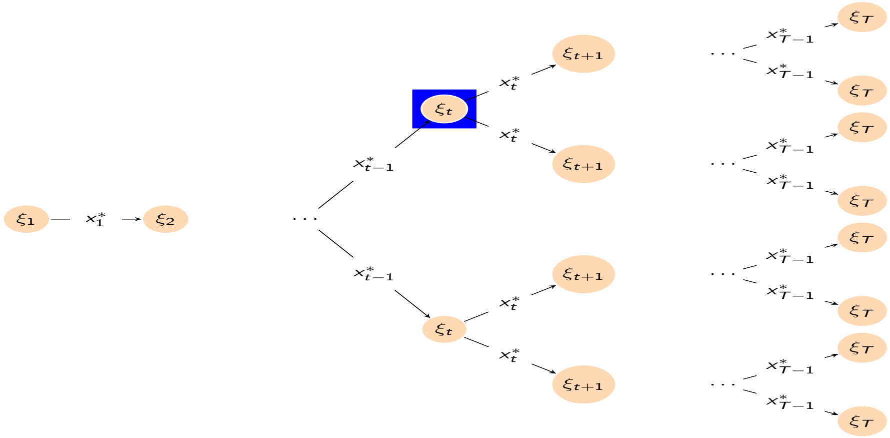
Toward a Nested Formulation of Multistage DRSP

Given a scenario tree and a nominal distribution on the tree



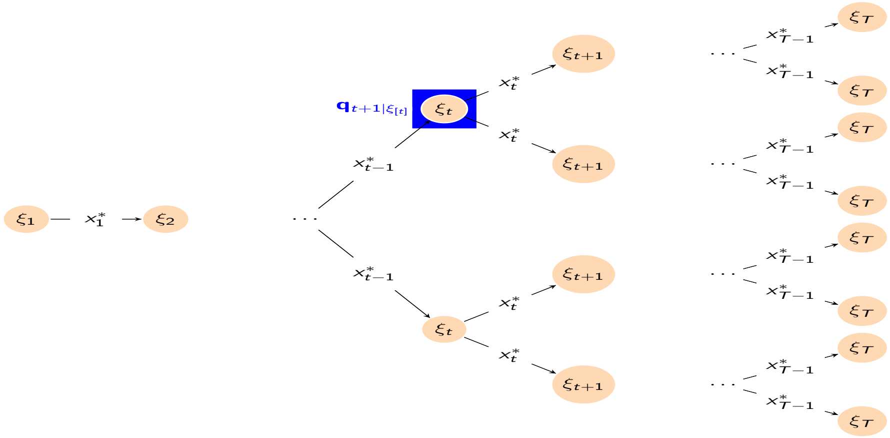
Toward a Nested Formulation of Multistage DRSP

Given a scenario tree and a nominal distribution on the tree



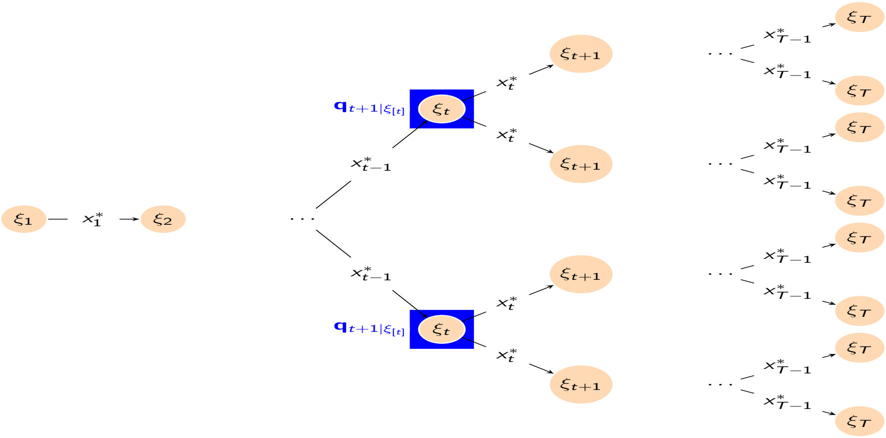
Toward a Nested Formulation of Multistage DRSP

Given a scenario tree and a nominal distribution on the tree



Toward a Nested Formulation of Multistage DRSP

Given a scenario tree and a nominal distribution on the tree



Nested Formulation of Multistage DRSP

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \mathbb{E}_{\mathbf{q}_2|\xi_{[1]}} \left[\min_{x_2 \in \mathcal{X}_2} g_2(x_2, \xi_2) + \mathbb{E}_{\mathbf{q}_3|\xi_{[2]}} \left[\cdots + \mathbb{E}_{\mathbf{q}_T|\xi_{[T-1]}} \left[\min_{x_T \in \mathcal{X}_T} g_T(x_T, \xi_T) \right] \cdots \right] \right],$$

Nested Formulation of Multistage DRSP

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \max_{\mathbf{p}_2 \in \mathcal{P}_{2|\xi_{[1]}}} \mathbb{E}_{\mathbf{p}_2} \left[\min_{x_2 \in \mathcal{X}_2} g_2(x_2, \xi_2) + \max_{\mathbf{p}_3 \in \mathcal{P}_{3|\xi_{[2]}}} \mathbb{E}_{\mathbf{p}_3} \left[\cdots + \max_{\mathbf{p}_T \in \mathcal{P}_{T|\xi_{[T-1]}}} \mathbb{E}_{\mathbf{p}_T} \left[\min_{x_T \in \mathcal{X}_T} g_T(x_T, \xi_T) \right] \cdots \right] \right],$$

Nested Formulation of Multistage DRSP

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \max_{\mathbf{p}_2 \in \mathcal{P}_{2|\xi_{[1]}}} \mathbb{E}_{\mathbf{p}_2} \Bigg[\min_{x_2 \in \mathcal{X}_2} g_2(x_2, \xi_2) + \max_{\mathbf{p}_3 \in \mathcal{P}_{3|\xi_{[2]}}} \mathbb{E}_{\mathbf{p}_3} \Big[\dots + \max_{\mathbf{p}_T \in \mathcal{P}_{T|\xi_{[T-1]}}} \mathbb{E}_{\mathbf{p}_T} \Big[\min_{x_T \in \mathcal{X}_T} g_T(x_T, \xi_T) \Big] \dots \Big] \Bigg],$$

where

$\mathcal{P}_{t|\xi_{[t-1]}}$ is the conditional ambiguity set for stage-t probability measure, conditioned on $\xi_{[t-1]}$.

How to Construct the Ambiguity Set (Multistage)?

- *Moment*-based sets: distributions with similar moments

  (Shapiro, 2012), (Xin et al., 2013), (Xin and Goldberg, 2015)

- *Distance*-based sets: sufficiently close distributions to a nominal distribution with respect to a distance
  - *Nested distance (Wasserstein metric):* (Pflug and Pichler, 2014), (Analui and Pflug, 2014)
  - *Modified* χ^2 *distance:* (Philpott et al., 2017)
  - L_∞ *norm:* (Huang et al., 2017)
  - *General theory:* (Shapiro, 2016; 2017; 2018)

How to Construct the Ambiguity Set (Multistage)?

- *Moment*-based sets: distributions with similar moments

  (Shapiro, 2012), (Xin et al., 2013), (Xin and Goldberg, 2015)

- *Distance*-based sets: sufficiently close distributions to a nominal distribution with respect to a distance
  - *Nested distance (Wasserstein metric):* (Pflug and Pichler, 2014), (Analui and Pflug, 2014)
  - *Modified* χ^2 *distance:* (Philpott et al., 2017)
  - L_∞ *norm:* (Huang et al., 2017)
  - *General theory:* (Shapiro, 2016; 2017; 2018)
  - **Total variation distance**

Multistage DRSP with Total Variation Distance (DRSP-V)

At stage t, given $\xi_{[t-1]}$, instead of considering one ("nominal") distribution $\mathbf{q}_{t|\xi_{[t-1]}}$,

Consider all distributions $\mathbf{p}_t$ in

$$\mathcal{P}_{t|\xi_{[t-1]}} = \left\{ \mathbf{p}_t : \mathsf{V}(\mathbf{p}_t, \mathbf{q}_{t|\xi_{[t-1]}}) := \frac{1}{2} \int_{\Xi_{t|\xi_{[t-1]}}} \left| \mathbf{p}_t - \mathbf{q}_{t|\xi_{[t-1]}} \right| \, d\nu \leq \gamma_t, \right.$$
$$\int_{\Xi_{t|\xi_{[t-1]}}} \mathbf{p}_t \, d\nu = 1,$$
$$\left. \mathbf{p}_t \geq 0 \right\},$$

where $\Xi_{t|\xi_{[t-1]}}$ is the sample space of stage t, given $\xi_{[t-1]}$.

Multistage DRSP with Total Variation Distance (DRSP-V)

At stage t, given $\xi_{[t-1]}$, instead of considering one ("nominal") distribution $\mathbf{q}_{t|\xi_{[t-1]}}$,

Consider all distributions $\mathbf{p}_t$ in

$$\mathcal{P}_{t|\xi_{[t-1]}} = \left\{ \mathbf{p}_t : \mathrm{V}(\mathbf{p}_t, \mathbf{q}_{t|\xi_{[t-1]}}) := \frac{1}{2} \int_{\Xi_{t|\xi_{[t-1]}}} \left| \mathbf{p}_t - \mathbf{q}_{t|\xi_{[t-1]}} \right| \, d\nu \leq \gamma_t, \right.$$
$$\int_{\Xi_{t|\xi_{[t-1]}}} \mathbf{p}_t \, d\nu = 1,$$
$$\left. \mathbf{p}_t \geq 0 \right\},$$

where $\Xi_{t|\xi_{[t-1]}}$ is the sample space of stage t, given $\xi_{[t-1]}$.

- all distributions sufficiently close to the nominal distribution

Multistage DRSP with Total Variation Distance (DRSP-V)

At stage t, given $\xi_{[t-1]}$, instead of considering one ("nominal") distribution $\mathbf{q}_{t|\xi_{[t-1]}}$,

Consider all distributions $\mathbf{p}_t$ in

$$\mathcal{P}_{t|\xi_{[t-1]}} = \Bigg\{ \mathbf{p}_t : \mathrm{V}(\mathbf{p}_t, \mathbf{q}_{t|\xi_{[t-1]}}) := \frac{1}{2} \int_{\Xi_{t|\xi_{[t-1]}}} \left| \mathbf{p}_t - \mathbf{q}_{t|\xi_{[t-1]}} \right| \, d\nu \leq \gamma_t, \\ \int_{\Xi_{t|\xi_{[t-1]}}} \mathbf{p}_t \, d\nu = 1, \\ \mathbf{p}_t \geq 0 \Bigg\},$$

where $\Xi_{t|\xi_{[t-1]}}$ is the sample space of stage t, given $\xi_{[t-1]}$.

- ensure it is a probability measure

Aim

Q1: How do we formulate this problem?

Q2: What uncertain scenarios are *important* to a multistage DRSP model?

- How to define *important* scenarios?
- How to identify *important* scenarios?

But ... Let's take a look at **static/two-stage** case first

Outline

Static/Two-Stage DRSP

$$\min_{x \in \mathcal{X}} \left\{ f(x) := \max_{\mathbf{p} \in \mathcal{P}} \mathbb{E}_{\mathbf{p}} \left[h(x, \omega) \right] \right\},$$

where

- $\mathcal{X} \subseteq \mathbb{R}^n$ is a deterministic and non-empty convex compact set,
- Ω is sample space, assumed finite
- $h : \mathcal{X} \times \Omega \mapsto \mathbb{R}$ is an integrable convex random function, i.e., for any $x \in \mathcal{X}$, $h(x, \cdot)$ is integrable, and $h(\cdot, \omega)$ is convex q-almost surely,

Static/Two-Stage DRSP

$$\min_{x\in\mathcal{X}} \left\{ f(x) := \max_{\mathbf{p}\in\mathcal{P}} \mathbb{E}_{\mathbf{p}}\left[h(x,\omega)\right] \right\},$$

where

- $\mathbf{q}$ denotes a nominal probability distribution, which may be obtained from data, e.g., empirical distribution,
- $\mathcal{P}$ is the ambiguity set of distributions, a subset of all probability distributions on Ω, which may be obtained, e.g., via the total variation distance to the nominal distribution

Assessment Problem of "Removed" Scenarios

Consider "removing" a set $\mathcal{F} \subset \Omega$ of scenarios:

$$\mathcal{P}^{\mathsf{A}} := \{\mathbf{p} \in \mathcal{P} : p_\omega = 0,\ \omega \in \mathcal{F}\}.$$

Assessment Problem of "Removed" Scenarios

Consider "removing" a set $\mathcal{F} \subset \Omega$ of scenarios:

$$\mathcal{P}^{\mathsf{A}} := \{\mathbf{p} \in \mathcal{P} : p_\omega = 0,\ \omega \in \mathcal{F}\}.$$

Assessment Problem of "Removed" Scenarios

Consider "removing" a set $\mathcal{F} \subset \Omega$ of scenarios:

$$\mathcal{P}^{\mathsf{A}} := \{\mathbf{p} \in \mathcal{P} : p_\omega = 0,\ \omega \in \mathcal{F}\}.$$

The Assessment problem of scenarios in $\mathcal{F}$ is

$$\min_{x \in \mathbb{X}} \left\{ f^{\mathsf{A}}(x; \mathcal{F}) = \max_{\mathbf{p} \in \mathcal{P}^{\mathsf{A}}(\mathcal{F})} \sum_{\omega \in \mathcal{F}^{\mathsf{c}}} p_\omega h_\omega(x) \right\},$$

where

If Inner Max of the Assessment Problem is Infeasible: $f^{\mathsf{A}}(x; \mathcal{F}) = -\infty$

Assessment Problem of "Removed" Scenarios

Consider "removing" a set $\mathcal{F} \subset \Omega$ of scenarios:

$$\mathcal{P}^{\mathsf{A}} := \{\mathbf{p} \in \mathcal{P} : p_\omega = 0,\ \omega \in \mathcal{F}\}.$$

The Assessment problem of scenarios in $\mathcal{F}$ is

$$\min_{x \in \mathbb{X}} \left\{ f^{\mathsf{A}}(x; \mathcal{F}) = \max_{\mathbf{p} \in \mathcal{P}^{\mathsf{A}}(\mathcal{F})} \sum_{\omega \in \mathcal{F}^{\mathsf{c}}} p_\omega h_\omega(x) \right\},$$

where

If Inner Max of the Assessment Problem is Infeasible: $f^{\mathsf{A}}(x; \mathcal{F}) = -\infty$

Assessment Problem of "Removed" Scenarios

Consider "removing" a set $\mathcal{F} \subset \Omega$ of scenarios:

$$\mathcal{P}^{\mathsf{A}} := \{\mathbf{p} \in \mathcal{P} : p_\omega = 0,\ \omega \in \mathcal{F}\}.$$

The Assessment problem of scenarios in $\mathcal{F}$ is

$$\min_{x \in \mathbb{X}} \left\{ f^{\mathsf{A}}(x; \mathcal{F}) = \max_{\mathbf{p} \in \mathcal{P}^{\mathsf{A}}(\mathcal{F})} \sum_{\omega \in \mathcal{F}^{\mathsf{c}}} p_\omega h_\omega(x) \right\},$$

where

If Inner Max of the Assessment Problem is Infeasible: $f^{\mathsf{A}}(x; \mathcal{F}) = -\infty$

Effective/Ineffective Scenarios in DRSP

(Rahimian, B., Homem-de-Mello, 2018)

Definition (Effective Subset of Scenarios)

At an optimal solution x^*, a subset $\mathcal{F} \subset \Omega$ is called effective if by its "removal" the optimal value of the Assessment problem is strictly smaller than the optimal value of DRSP; i.e., if

$$\min_{x \in \mathcal{X}} f^{\mathsf{A}}(x; \mathcal{F}) < \min_{x \in \mathcal{X}} f(x)$$

.

Effective/Ineffective Scenarios in DRSP

(Rahimian, B., Homem-de-Mello, 2018)

Definition (**Effective Subset of Scenarios)**

At an optimal solution x^*, a subset $\mathcal{F} \subset \Omega$ is called effective if by its "removal" the optimal value of the Assessment problem is strictly smaller than the optimal value of DRSP; i.e., if

$$\min_{x \in \mathcal{X}} f^{\mathsf{A}}(x; \mathcal{F}) < \min_{x \in \mathcal{X}} f(x)$$

.

Definition (**Ineffective Subset of Scenarios)**

A subset $\mathcal{F} \subset \Omega$ that is not effective is called ineffective.

DRSP with Total Variation Distance

$$\min_{x \in \mathcal{X}} \max_{p \in \mathcal{P}} \sum_{\omega=1}^{n} p_\omega h(x, \omega)$$

where

$$\mathcal{P} = \left\{ \frac{1}{2} \sum_{\omega \in \Omega} |p_\omega - q_\omega| \leq \gamma, \ \sum_{\omega=1}^{n} p_\omega = 1, p_\omega \geq 0, \forall \omega \right\},$$

Risk-Averse Interpretation

Proposition (**Risk-Averse Interpretation of DRSP with Total Variation**)

$$f_\gamma(x) = \begin{cases} \mathbb{E}_{\mathbf{q}}\left[h(x,\omega)\right], & \text{if } \gamma = 0, \\ \gamma \sup_{\omega\in\Omega} h(x,\omega) + (1-\gamma)\,\mathrm{CVaR}_\gamma\left[h(x,\omega)\right], & \text{if } 0 < \gamma < 1, \\ \sup\limits_{\omega\in\Omega} h(x,\omega), & \text{if } \gamma \geq 1, \end{cases}$$

By (Jiang and Guan, 2016).

How to Find Effective/Ineffective Scenarios for DRSP?

How can we determine the effectiveness of a scenario?

- Resolve for any scenario $\omega \in \Omega$
  - Form the corresponding Assessment problem,
  - Resolve the corresponding Assessment problem,
  - Compare the optimal values to determine the effectiveness of the scenario.
- Exploit the structure of the ambiguity set
  - Propose easy-to-check conditions (based on optimal solution and worst-case distribution) to identify the effectiveness of a scenario
  - Low computational cost
  - We might not be able to identify the effectiveness of all scenarios

Notation

Consider an optimal solution $(x^*, \mathbf{p}^*) \in \mathcal{X} \times \mathcal{P}$ to DRSP-V:

$$x^* \in \operatorname*{argmin}_{x \in \mathcal{X}} \mathbb{E}_{\mathbf{p}*}\left[h(x,\omega)\right]$$
$$\mathbf{p}^* := \mathbf{p}^*(x^*) \in \operatorname*{argmax}_{\mathbf{p} \in \mathcal{P}} \mathbb{E}_{\mathbf{p}}\left[h(x^*,\omega)\right]$$

Define

$$\Omega_1(x^*) := \left[\omega \in \Omega : h(x^*,\omega) < \mathrm{VaR}_\gamma\left[h(x^*,\omega)\right]\right]$$
$$\Omega_2(x^*) := \left[\omega \in \Omega : h(x^*,\omega) = \mathrm{VaR}_\gamma\left[h(x^*,\omega)\right]\right]$$
$$\Omega_3(x^*) := \left[\omega \in \Omega : \mathrm{VaR}_\gamma\left[h(x^*,\omega)\right] < h(x^*,\omega) < \sup_{\omega \in \Omega} h(x^*,\omega)\right]$$
$$\Omega_4(x^*) := \left[\omega \in \Omega : h(x^*,\omega) = \sup_{\omega \in \Omega} h(x^*,\omega)\right]$$

Ineffective Scenarios

Theorem (**Easy-to-Check Conditions for Ineffective Scenarios,**
(Rahimian, B., Homem-de-Mello, 2018))

Suppose $(x^*, \mathbf{p}^*)$ *solves DRSP-V. Then, a scenario* ω' *with* $q_{\omega'} \leq \gamma$, *is ineffective if any of the following conditions holds:*

- $\omega' \in \Omega_1(x^*)$,
- $\omega' \in \Omega_2(x^*)$ *and* $q_{\omega'} = 0$,
- $\omega' \in \Omega_2(x^*)$ *and* $\sum_{\omega \in \Omega_2(x^*)} p^*_\omega = 0$,
- $\omega' \in \Omega_3(x^*)$ *and* $q_{\omega'} = 0$.

Effective Scenarios

Theorem (Easy-to-Check Conditions for Effective Scenarios)

Suppose $(x^*, \mathbf{p}^*)$ solves DRSP-V. Then, a scenario ω' is effective if any of the following conditions holds:

- $q_{\omega'} > \gamma$,
- $\Omega_2(x^*) = \{\omega'\}$ *and* $p^*_{\omega'} > 0$,
- $\omega' \in \Omega_3(x^*)$ *and* $q_{\omega'} > 0$,
- $\omega' \in \Omega_4(x^*)$ *and* $q_{\omega'} > 0$,
- $\Omega_4(x^*) = \{\omega'\}$.

Effective Scenarios

Theorem (Easy-to-Check Conditions for Effective Scenarios)

Suppose $(x^*, \mathbf{p}^*)$ solves DRSP-V. Then, a scenario ω' is effective if any of the following conditions holds:

- $q_{\omega'} > \gamma$,
- $\Omega_2(x^*) = \{\omega'\}$ *and* $p^*_{\omega'} > 0$,
- $\omega' \in \Omega_3(x^*)$ *and* $q_{\omega'} > 0$,
- $\omega' \in \Omega_4(x^*)$ *and* $q_{\omega'} > 0$,
- $\Omega_4(x^*) = \{\omega'\}$.

▶ Trivially Effective !

Beyond Previous Theorems: Identify Undetermined Scenarios

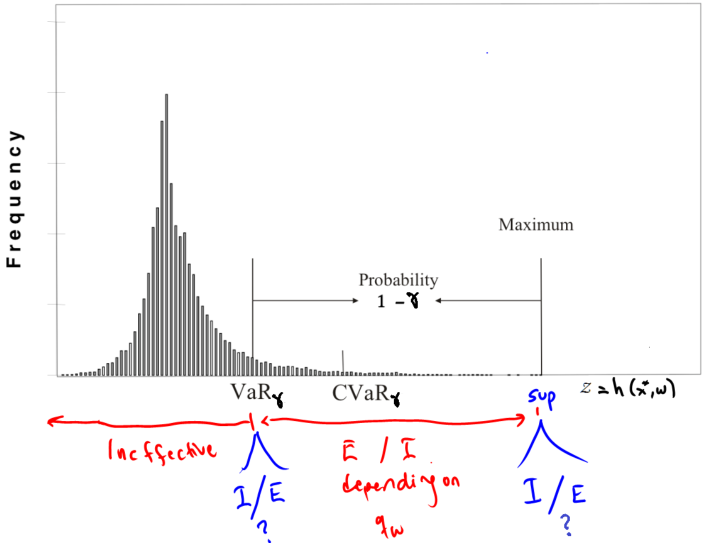
Theorem (Easy-to-Check Conditions to Identify Undetermined Scenarios)

Suppose $(x^*, \mathbf{p}^*)$ solves DRO-V. For a scenario $\omega' \in \Omega_2(x^*)$ with $q_{\omega'} > 0$, suppose that the effectiveness of scenario ω' is not identified by the previous theorems. Let $\mathcal{F} = \{\omega'\}$. If

1. $\mathrm{VaR}_{\gamma_{\mathcal{F}}}\left[h(x^*, \omega) | \mathcal{F}^c\right] < \mathrm{VaR}_{\gamma}\left[h(x^*, \omega)\right]$, *and*
2. *either there exists a scenario* $\omega \in \left[\mathrm{VaR}_{\gamma_{\mathcal{F}}}\left[h(x^*, \omega) | \mathcal{F}^c\right] < h(x^*, \omega) < \mathrm{VaR}_{\gamma}\left[h(x^*, \omega)\right]\right]$ *with* $q_\omega > 0$ *or* $\Psi_{|\mathcal{F}^c}\left(x^*, \mathrm{VaR}_{\gamma_{\mathcal{F}}}\left[h(x^* \omega), |\mathcal{F}^c\right]\right) > \gamma_{\mathcal{F}}$,

then scenario ω' is effective.

Effective/Ineffective Scenarios Summary (Two-Stage)



Outline

Effective/Ineffective Scenarios in Multistage DRSP

What happens in the Multistage case?

Relation to Multistage Risk-Averse Optimization

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \max_{\mathbf{p}_2 \in \mathcal{P}_{2|\xi_{[1]}}} \mathbb{E}_{\mathbf{p}_2} \Bigg[\min_{x_2 \in \mathcal{X}_2} g_2(x_2, \xi_2) + \max_{\mathbf{p}_3 \in \mathcal{P}_{3|\xi_{[2]}}} \dots$$
$$\dots + \max_{\mathbf{p}_T \in \mathcal{P}_{T|\xi_{[T-1]}}} \mathbb{E}_{\mathbf{p}_T} \Bigg[\min_{x_T \in \mathcal{X}_T} g_T(x_T, \xi_T) \Bigg] \dots \Bigg] \Bigg]$$

Relation to Multistage Risk-Averse Optimization

Proposition (**Risk-Averse Interpretation of Multistage DRSP-V)**

Multistage DRSP-V can be written as

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \mathcal{R}_{2|\xi_{[1]}} \left[\min_{x_2 \in \mathcal{X}_2} g_2(x_2, \xi_2) + \mathcal{R}_{3|\xi_{[2]}} \left[\ldots + \mathcal{R}_{T|\xi_{[T-1]}} \left[\min_{x_T \in \mathcal{X}_T} g_T(x_T, \xi_T) \right] \ldots \right] \right],$$

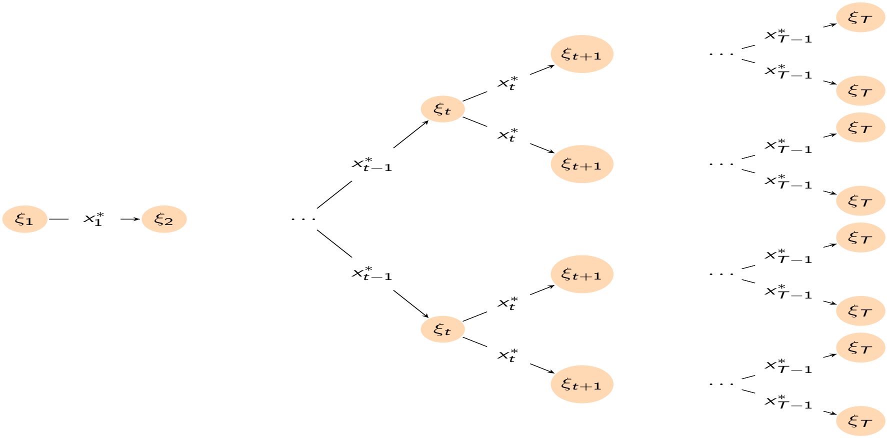
where $\mathcal{R}$'s are the (real-valued) coherent conditional risk mappings

$$\mathcal{R}_{t+1|\xi_{[t]}} [\cdot] = \begin{cases} \mathbb{E}_{\mathbf{q}_{\mathbf{t+1}|\xi_{[\mathbf{t}]}}} [\cdot], & \text{if } \gamma = 0, \\ \gamma \sup_{\xi_{t+1} \in \Xi_{t+1|\xi_{[t]}}} [\cdot] + (1-\gamma) \, \mathrm{CVaR}_\gamma [\cdot], & \text{if } 0 < \gamma < 1, \\ \sup_{\xi_{t+1} \in \Xi_{t+1|\xi_{[t]}}} [\cdot], & \text{if } \gamma \geq 1. \end{cases}$$

where $\cdot$ is $Q_{t+1}(x_{[t]}, \xi_{[t+1]})$ is the cost-to-go function.

Effective/Ineffective Scenarios in Multistage DRSP?

Now we have a scenario tree. What to do?



Effective/Ineffective Scenarios in Multistage DRSP?

Questions

- What is the effectiveness of a scenario (path)?
- What is the effectiveness of a realization in stage $t+1$?

Effective/Ineffective Scenarios in Multistage DRSP?

Questions

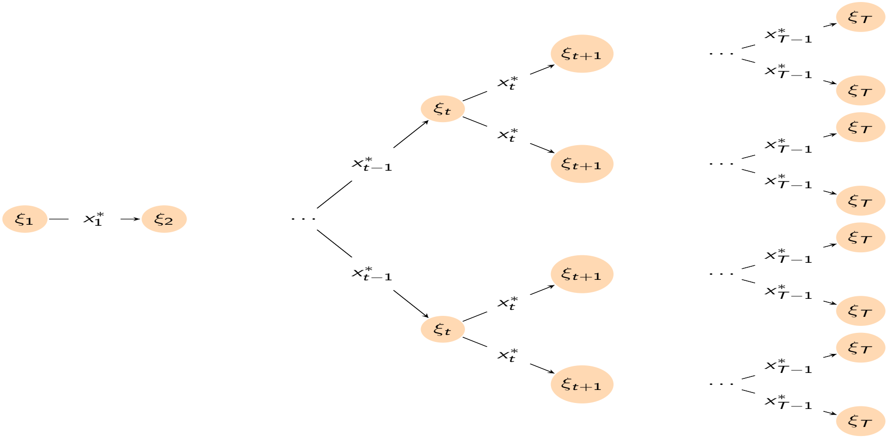
- What is the effectiveness of a scenario (path)?
- What is the effectiveness of a realization in stage $t+1$?

Main Idea

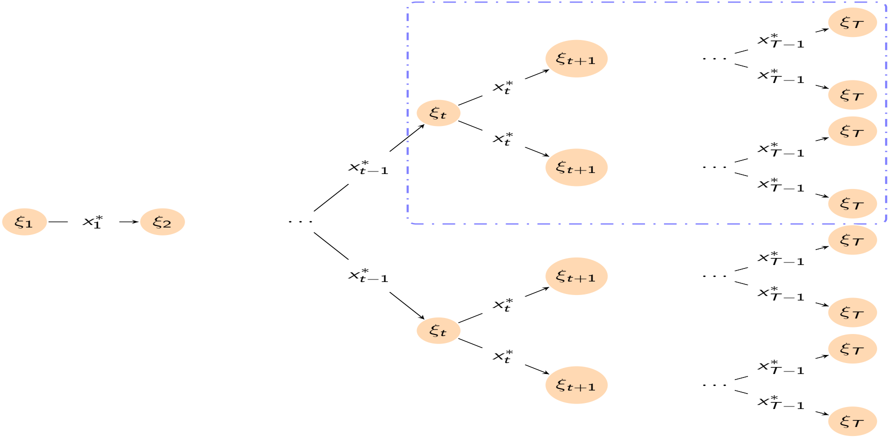
- Look at realizations conditioned on their history of decisions and stochastic process

$\rightarrow$ At an optimal policy x^*, if we look at stage t, **given** $x^*_{[t-1]}$ and $\xi_{[t]}$, previous definitions on effective/ineffective scenarios hold **conditionally**.

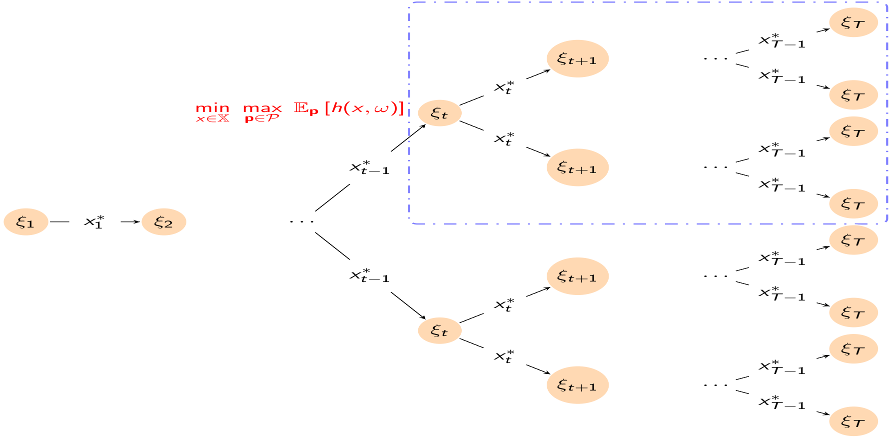
Effective/Ineffective Scenarios in Multistage DRSP?



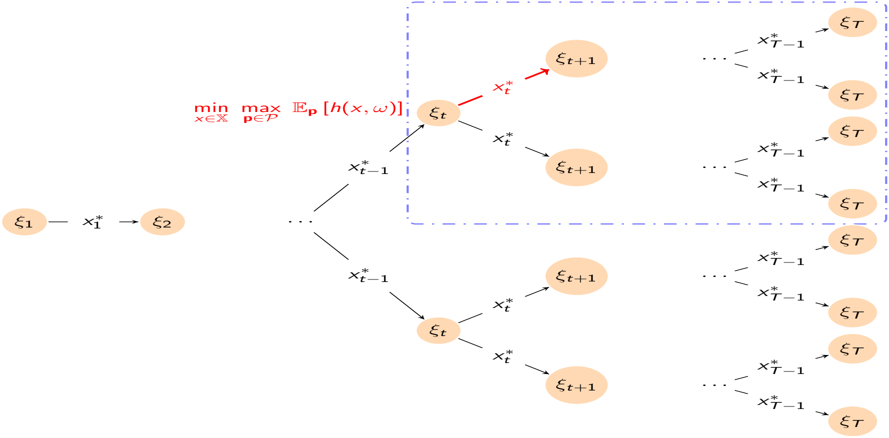
Effective/Ineffective Scenarios in Multistage DRSP?



Effective/Ineffective Scenarios in Multistage DRSP?



Effective/Ineffective Scenarios in Multistage DRSP?



Effective Scenarios in Multistage DRSP:

Conditional Effectiveness

Definition (Conditionally Effective Realization)

At an optimal policy $x^* := [x_1^*, \ldots, x_T^*]$, a realization of ξ_{t+1} in stage $t+1$ is called conditionally effective, given $x_{[t-1]}^*$ and $\xi_{[t]}$, if by its removal the optimal stage-t cost function (immediate cost + cost-to-go function) of the new problem is strictly smaller than the optimal value of the original stage-t problem in multistage DRSP.

Effective Scenarios in Multistage DRSP:

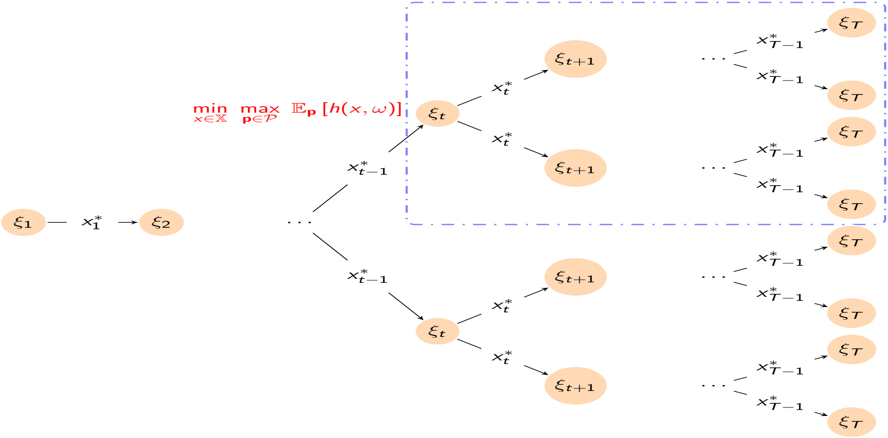
Effectiveness of a Scenario Path

Definition (**Effective Scenario Path**)

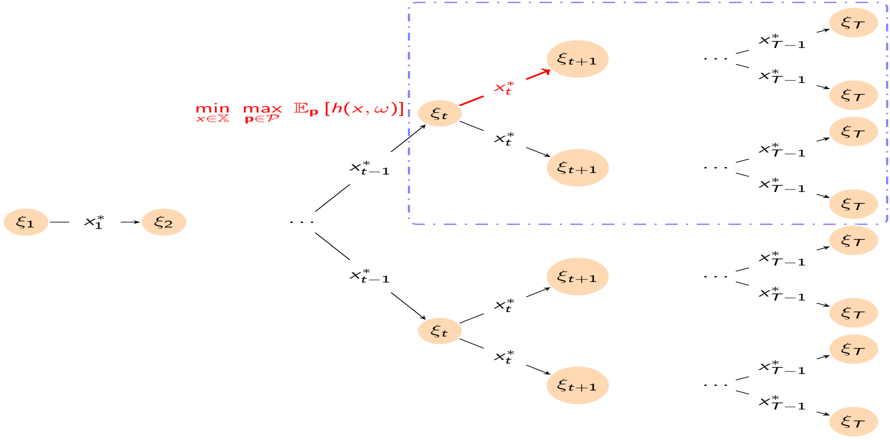
At an optimal policy $x^* := [x_1^*, \dots, x_T^*]$, a scenario path $\{\xi_t\}_{t=1}^T$ is called effective if by its "removal" the optimal value of the new problem is strictly smaller than the optimal value of multistage DRSP.

NOTE: Removing a scenario path is defined by forcing the probability of ξ_T to be zero.

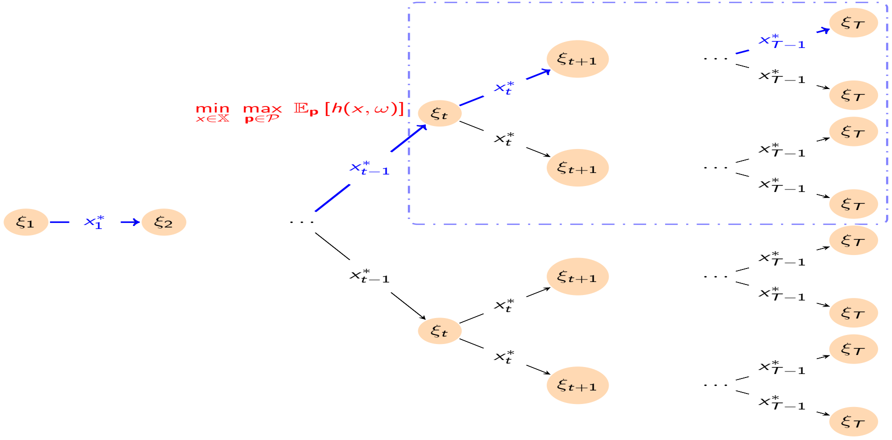
Difference Between Conditional Effective Realizations and Effective Scenario Paths



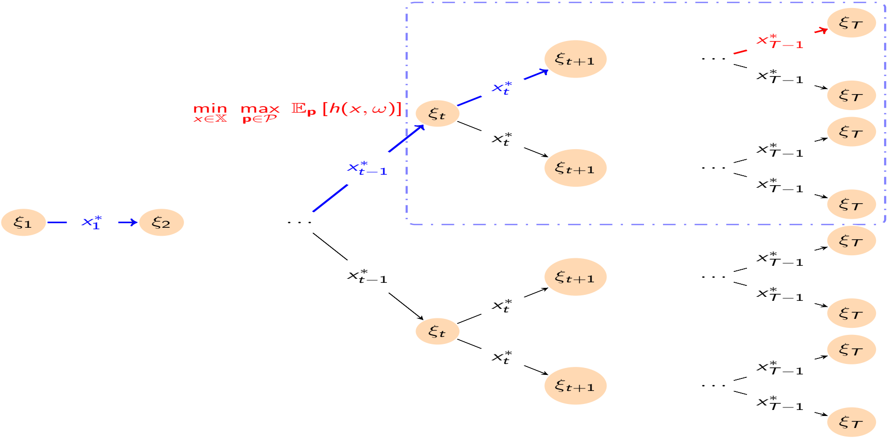
Difference Between Conditional Effective Realizations and Effective Scenario Paths



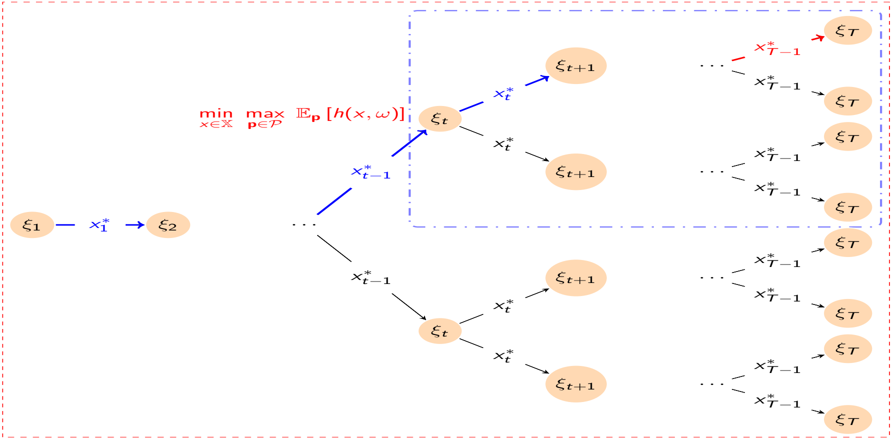
Difference Between Conditional Effective Realizations and Effective Scenario Paths



Difference Between Conditional Effective Realizations and Effective Scenario Paths



Difference Between Conditional Effective Realizations and Effective Scenario Paths



How to Find Effective/Ineffective Scenarios for Multistage DRSP-V?

How to Find Effective/Ineffective Scenarios for Multistage DRSP-V?

Resolve?

How to Find Effective/Ineffective Scenarios for Multistage DRSP-V?

Resolve?

Suppose each node has n children. Then, we would have to solve many problems!

- **Effectiveness of Scenario Paths**: n^{T-1} problems at stage T

How to Find Effective/Ineffective Scenarios for Multistage DRSP-V?

Resolve?

Suppose each node has n children. Then, we would have to solve many problems!

- **Effectiveness of Scenario Paths**: n^{T-1} problems at stage T
- **Conditionally Effectiveness of Realizations**: $n + \ldots + n^{T-1}$ problems at stage $2 + \ldots +$ stage T

How to Find Effective/Ineffective Scenarios for Multistage DRSP-V?

Resolve?

Suppose each node has n children. Then, we would have to solve many problems!

- **Effectiveness of Scenario Paths**: n^{T-1} problems at stage T
- **Conditionally Effectiveness of Realizations**: $n + \ldots + n^{T-1}$ problems at stage 2 $+ \ldots +$ stage T

$\rightarrow$ AIM: Propose easy-to-check conditions

Use Conditional Effectiveness of Realizations in Multistage DRSP-V

AIM: Propose easy-to-check conditions

Theorem [Conditionally Multistage ← Two-stage]

Our easy-to-check conditions to identify effective/ineffective scenarios in static/two-stage DRSP-V are valid conditions to identify conditionally effective/ineffective scenarios in multistage DRSP-V.

Effectiveness of Scenario Paths in Multistage DRSP-V

Consider a scenario path $\{\xi_t\}_{t=1}^{T}$.

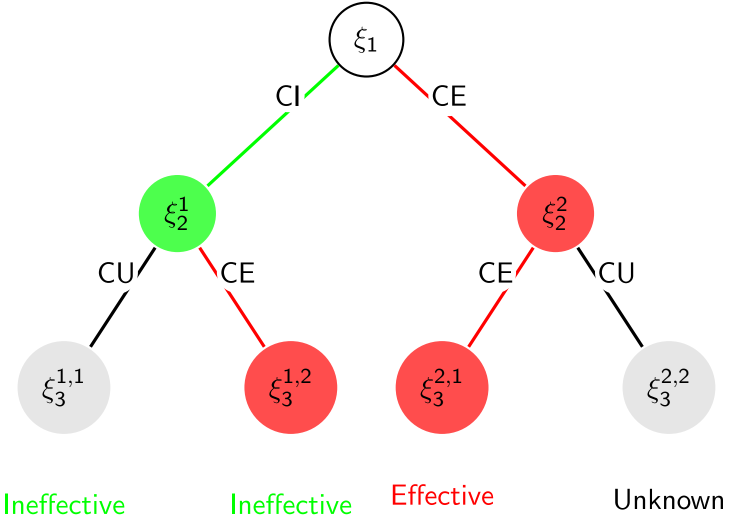
Theorem

If ξ_t is conditionally effective by our easy-to-check conditions, for all $t = 1, \ldots, T$, then, the scenario path $\{\xi_t\}_{t=1}^{T}$ is effective.

Theorem

If ξ_T is **not trivially** conditionally effective (i.e., too large nominal conditional probability) and there exists t, $t = 1, \ldots, T$, such that ξ_t is conditionally ineffective by our easy-to-check conditions, then, the scenario path $\{\xi_t\}_{t=1}^{T}$ is ineffective.

Easy-To-Check Conditions for Effectiveness of Scenario Paths



Outline

Dynamic Programming Formulation

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \max_{\mathbf{p}_2 \in \mathcal{P}_{2|\xi_{[1]}}} \mathbb{E}_{\mathbf{p}_2} \left[\min_{x_2 \in \mathcal{X}_2} g_2(x_2, \xi_2) + \ldots + \max_{\mathbf{p}_T \in \mathcal{P}_{T|\xi_{[T-1]}}} \mathbb{E}_{\mathbf{p}_T} \left[\min_{x_T \in \mathcal{X}_T} g_T(x_T, \xi_T) \right] \right]$$

Dynamic Programming Formulation

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \max_{\mathbf{p}_2 \in \mathcal{P}_{2|\xi_{[1]}}} \mathbb{E}_{\mathbf{p}_2} \underbrace{\left[\min_{x_2 \in \mathcal{X}_2} g_2(x_2, \xi_2) + \ldots + \max_{\mathbf{p}_T \in \mathcal{P}_{T|\xi_{[T-1]}}} \mathbb{E}_{\mathbf{p}_T} \left[\min_{x_T \in \mathcal{X}_T} g_T(x_T, \xi_T) \right] \right]}_{Q_2(x_1, \xi_{[2]})}$$

First-stage cost function

$$\min_{x_1 \in \mathcal{X}_1} \; g_1(x_1, \xi_1) + \max_{\mathbf{p}_2 \in \mathcal{P}_{2|\xi_{[1]}}} \mathbb{E}_{\mathbf{p}_2} \left[Q_2(x_1, \xi_{[2]}) \right]$$

Dynamic Programming Formulation

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \max_{\mathbf{p}_2 \in \mathcal{P}_{2|\xi_{[1]}}} \mathbb{E}_{\mathbf{p}_2} \underbrace{\left[\min_{x_2 \in \mathcal{X}_2} g_2(x_2, \xi_2) + \underbrace{\ldots + \max_{\mathbf{p}_T \in \mathcal{P}_{T|\xi_{[T-1]}}} \mathbb{E}_{\mathbf{p}_T} \underbrace{\left[\min_{x_T \in \mathcal{X}_T} g_T(x_T, \xi_T) \right]}_{Q_T(x_{T-1}, \xi_{[T]})}}_{Q_3(x_2, \xi_{[3]})} \right]}_{Q_2(x_1, \xi_{[2]})}$$

First-stage cost function

$$\min_{x_1 \in \mathcal{X}_1} g_1(x_1, \xi_1) + \max_{\mathbf{p}_2 \in \mathcal{P}_{2|\xi_{[1]}}} \mathbb{E}_{\mathbf{p}_2} \left[Q_2(x_1, \xi_{[2]}) \right]$$

stage-t cost function

$$Q_t(x_{t-1}, \xi_{[t]}) := \min_{x_t \in \mathcal{X}_t} g_t(x_t, \xi_t) + \max_{\mathbf{p}_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}} \mathbb{E}_{\mathbf{p}_{t+1}} \left[Q_{t+1}(x_t, \xi_{[t+1]}) \right]$$

A Cutting Plane Approach

stage-t cost function

$$Q_t(x_{t-1}, \xi_{[t]}) = \min_{x_t \in \mathcal{X}_t} g_t(x_t, \xi_t) + \max_{\mathbf{p}_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}} \mathbb{E}_{\mathbf{p}_{t+1}} \left[Q_{t+1}(x_t, \xi_{[t+1]}) \right]$$

A Cutting Plane Approach

stage-t cost function

$$Q_t(x_{t-1}, \xi_{[t]}) = \min_{x_t \in \mathcal{X}_t} \; g_t(x_t, \xi_t) + \alpha_t$$

$$\text{s.t.} \quad \alpha_t \geq \max_{\mathbf{p}_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}} \mathbb{E}_{\mathbf{p}_{t+1}} \left[Q_{t+1}(x_t, \xi_{[t+1]}) \right]$$

A Cutting Plane Approach

stage-t cost function

$$Q_t(x_{t-1}, \xi_{[t]}) = \min_{x_t \in \mathcal{X}_t} \; g_t(x_t, \xi_t) + \alpha_t$$
$$\text{s.t.} \quad \alpha_t \geq \mathbb{E}_{\mathbf{p}_{t+1}} \left[Q_{t+1}(x_t, \xi_{[t+1]}) \right], \quad \mathbf{p}_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}$$

A Cutting Plane Approach

stage-t cost function

$$Q_t(x_{t-1}, \xi_{[t]}) = \min_{x_t \in \mathcal{X}_t} \; g_t(x_t, \xi_t) + \alpha_t$$
$$\text{s.t.} \quad \alpha_t \geq \mathbb{E}_{\mathbf{p}_{t+1}} \left[Q_{t+1}(x_t, \xi_{[t+1]}) \right], \quad \mathbf{p}_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}$$

For multistage DRSP-V,

- $\mathcal{P}_{t+1|\xi_{[t]}}$ is a polyhedron $\Longrightarrow$ Finite convergence

A Cutting Plane Approach

stage-t cost function

$$Q_t(x_{t-1}, \xi_{[t]}) = \min_{x_t \in \mathcal{X}_t} \; g_t(x_t, \xi_t) + \alpha_t$$

$$\text{s.t.} \quad \alpha_t \geq \mathbb{E}_{\mathbf{p}_{t+1}} \left[Q_{t+1}(x_t, \xi_{[t+1]}) \right], \quad \mathbf{p}_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}$$

For multistage DRSP-V,

- $\mathcal{P}_{t+1|\xi_{[t]}}$ is a polyhedron $\Longrightarrow$ Finite convergence

This idea can be applied to any polyhedral ambiguity set, with finite convergence guaranteed

How to Generate Distributional Cuts?

Distribution Separation Problem

For a fixed $x_t \in \mathcal{X}_t$, solve

$$\max_{p_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}} \mathbb{E}_{\mathbf{p}_{t+1}} \left[Q_{t+1}(x_t, \xi_{[t+1]}) \right]$$

How to Generate Distributional Cuts?

Distribution Separation Problem

For a fixed $x_t \in \mathcal{X}_t$, solve

$$\max_{p_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}} \int_{\Xi_{t+1|\xi_{[t]}}} \mathbf{p}_{t+1} Q_{t+1}(x_t, \cdot) \, d\,\nu$$

How to Generate Distributional Cuts?

Distribution Separation Problem

For a fixed $x_t \in \mathcal{X}_t$, solve

$$\max_{p_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}} \int_{\Xi_{t+1|\xi_{[t]}}} \mathbf{p}_{t+1} Q_{t+1}(x_t, \cdot)\, d\,\nu$$

For multistage DRSP-V,

- $\mathcal{P}_{t+1|\xi_{[t]}}$ is a polytope $\Longrightarrow$ Optimum is obtained at an extreme point

How to Generate Distributional Cuts?

Distribution Separation Problem

For a fixed $x_t \in \mathcal{X}_t$, solve

$$\max_{p_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}} \int_{\Xi_{t+1|\xi_{[t]}}} \mathbf{p}_{t+1} Q_{t+1}(x_t, \cdot) \, d\,\nu$$

For multistage DRSP-V,

- $\mathcal{P}_{t+1|\xi_{[t]}}$ is a polytope $\Longrightarrow$ Optimum is obtained at an extreme point

Challenge

- We do not have $Q_{t+1}(x_t, \xi_{[t+1]})$

How to Generate Distributional Cuts?

Distribution Separation Problem

For a fixed $x_t \in \mathcal{X}_t$, solve

$$\max_{p_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}} \int_{\Xi_{t+1|\xi_{[t]}}} \mathbf{p}_{t+1} \bar{Q}_{t+1}(x_t, \cdot)\, d\nu$$

For multistage DRSP-V,

- $\mathcal{P}_{t+1|\xi_{[t]}}$ is a polytope $\Longrightarrow$ Optimum is obtained at an extreme point

Challenge

- We do not have $Q_{t+1}(x_t, \xi_{[t+1]})$

But...

- We can use an inner (upper) approximation $\bar{Q}_{t+1}(x_t, \xi_{[t+1]})$

Primal Decomposition Algorithm

Main Idea

- Combine Nested L-shaped method and Distribution Separation problem

Forward Pass

- Obtain $x = [x_1, \ldots, x_T]$
- Use inner approximations on $Q_{t+1}(x_t, \xi_{[t+1]})$, $t = T-1, \ldots, 1$ to obtain $\mathbf{p} = [p_T, \ldots, p_2]$

Backward Pass

- Refine outer approximations on $Q_{t+1}(x_t, \xi_{[t+1]})$ and $\max_{\mathbf{p}_{t+1} \in \mathcal{P}_{t+1|\xi_{[t]}}} \mathbb{E}_{\mathbf{p}_{t+1}} \left[Q_{t+1}(x_t, \xi_{[t+1]}) \right]$

Outline

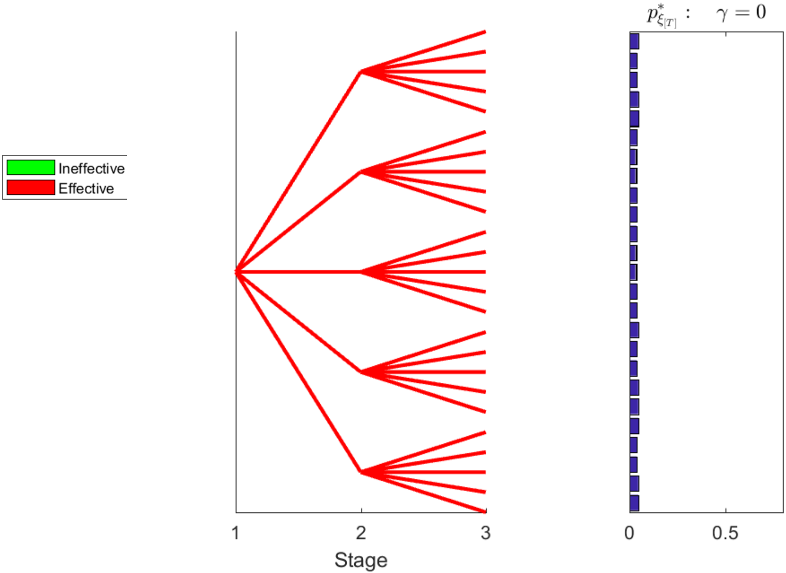
Test Problems

We considered two sets of problems:

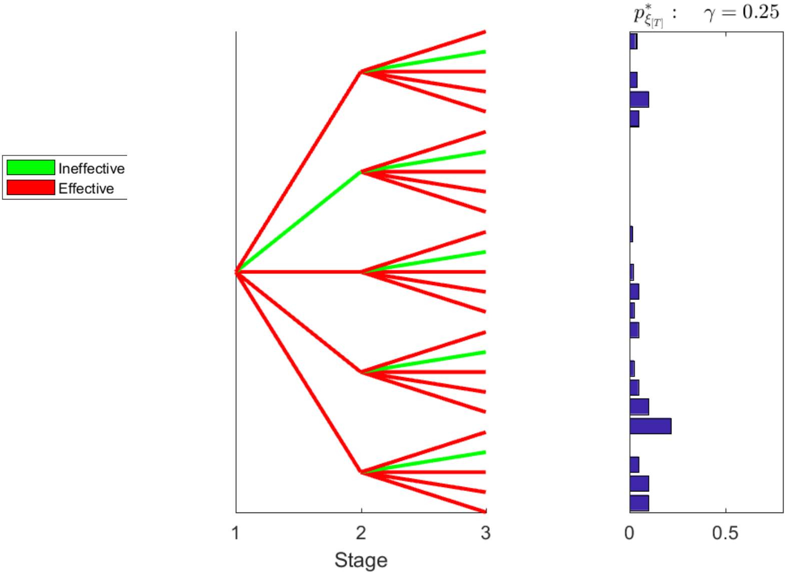
- SGPF—A Bond Investment Planning problem described by (Frauendorfer, Marohn, and SchÄurle, 1997) to maximize profit under uncertain returns
- Water Resources Allocation—Allocate Colorado River water among different users under water demand and supply uncertainties at minimum cost? (Zhang, Rahimian, Bayraksan, 2016)

We implemented our primal decomposition algorithm in C++ on top of SUTIL 0.1 (A Stochastic Programming Utility Library) (Czyzyk, Linderoth, and Shen, 2008) and solved problems with CPLEX 12.7.

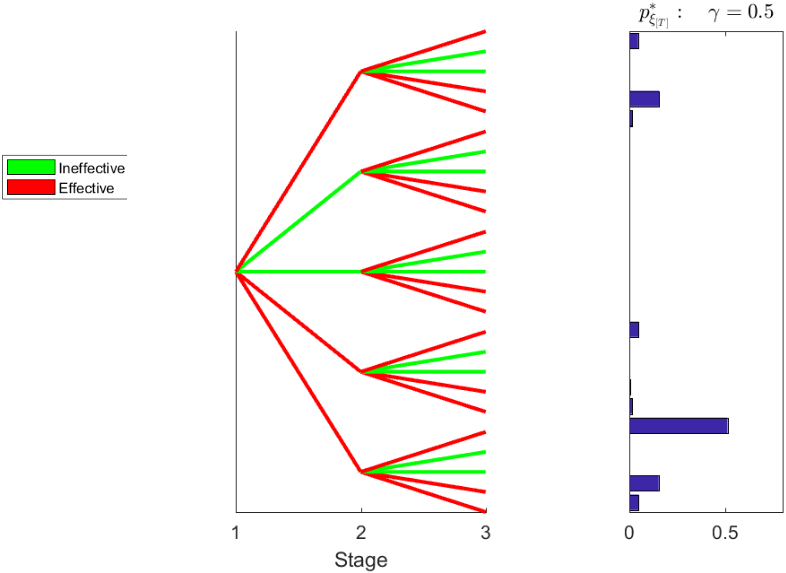
SGPF3Y3 (3 Stages, $5^2 = 25$ Scenarios)



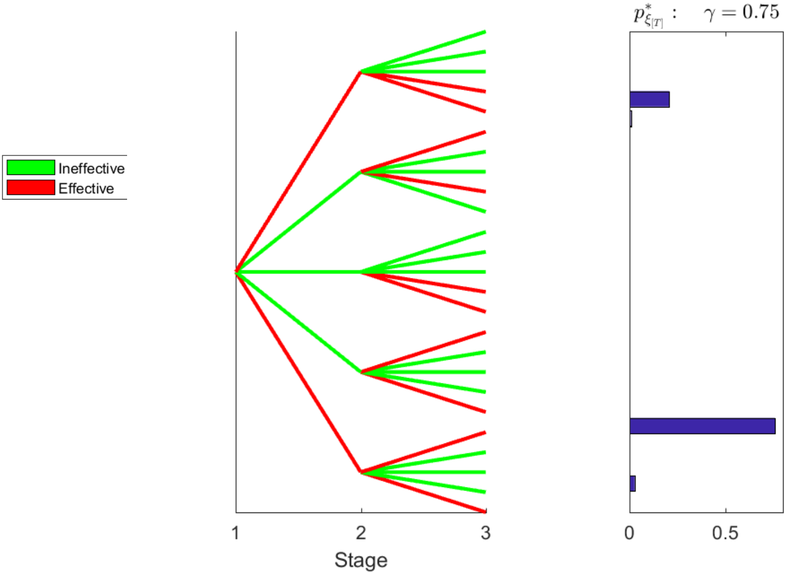
SGPF3Y3 (3 Stages, $5^2 = 25$ Scenarios)



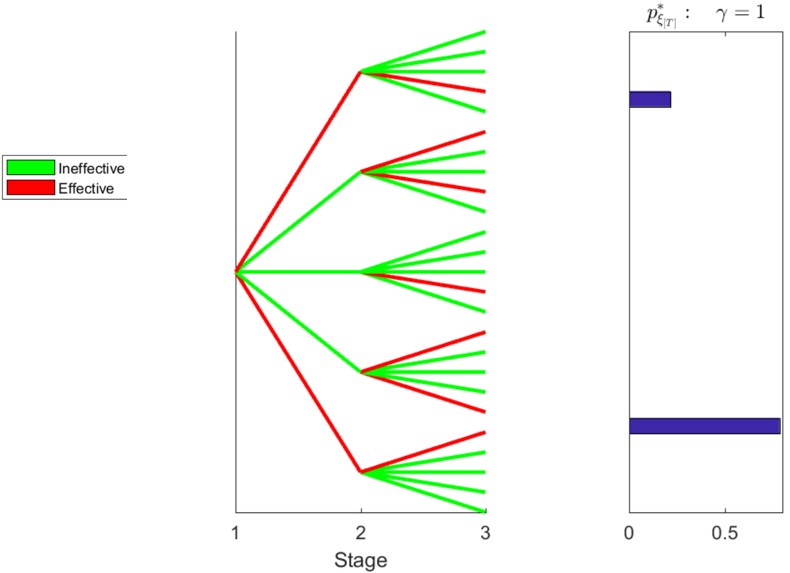
SGPF3Y3 (3 Stages, $5^2 = 25$ Scenarios)



SGPF3Y3 (3 Stages, $5^2 = 25$ Scenarios)



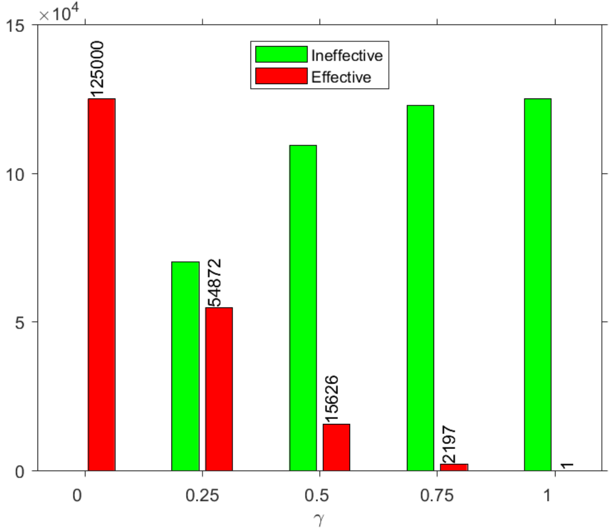
SGPF3Y3 (3 Stages, $5^2 = 25$ Scenarios)



SGPF3Y6 (6 Stages, $5^5 = 3125$ Scenarios)

	# of scenario path		
γ	ineffective	effective	undetermined
0.00	0	3125	0
0.05	0	3125	0
0.10	0	3125	0
0.15	0	3125	0
0.20	994	2131	0
0.25	2101	1024	0
0.30	2101	1024	0
0.35	2101	1024	0
0.40	2745	380	0
0.45	2793	183	149
0.50	2829	214	82
0.55	2873	234	18
0.60	3076	37	12
0.65	3081	24	20
0.70	3083	24	18
0.75	3089	36	0
0.80	3116	9	0
0.85	3116	9	0
0.90	3116	9	0
0.95	3116	9	0
1.00	3116	9	0

Water (4 Stages, $50^3 = 125 \times 10^3$ Scenarios)



Outline

Conclusion and Future Research

- Multistage DRSP-V is equivalent to a multistage risk-averse optimization, with a convex combination of worst-case and conditional value-at-risk as conditional risk mappings.

- Effective scenarios can provide managerial insight into the underlying uncertainties of the problems and encourage decision makers to collect more accurate information surrounding them.

- The notion of effective scenarios can be used for...
  - choosing the level of robustness
  - other ϕ-divergences and ambiguity sets
  - a better cut management in the primal decomposition algorithm
  - scenario reduction (Two-Stage: Aprón, Homem-de-Mello, Pagnocelli, 2018)
  - what happens as we add additional scenarios? Effective or not?

Acknowledgements and References

Gratefully acknowledge support of NSF through Grant CMMI-1563504 and DOE ASCR through Grant DE-AC02-06CH11347 (MACSER). Grateful to co-authors Hamed Rahimian and Tito Homem-de-Mello.

References:

- Rahimian, H., G. Bayraksan, and T. Homem-de-Mello, "Identifying Effective Scenarios in Distributionally Robust Stochastic Programs with Total Variation Distance," *Mathematical Programming*, 173(1-2): 393 – 430, 2019.
- Rahimian, H., G. Bayraksan, and T. Homem-de-Mello, "Distributionally Robust Newsvendor Problems with Variation Distance," Available at *Optimization Online*, 2017.
- Rahimian, H., G. Bayraksan, and T. Homem-de-Mello, "Effective Scenarios in Data-Driven Multistage Distributionally Robust Stochastic Programs with Total Variation Distance," *Working paper*.

Thank you!

(bayraksan.1@osu.edu)

DRSP with Total Variation Distance (DRSP-V)

Recall...

$$\min_{x \in \mathbb{X}} \left\{ f_\gamma(x) := \max_{\mathbf{p} \in \mathcal{P}} \sum_{\omega \in \Omega} p(\omega) h(x, \omega) \right\},$$

where

$$\mathcal{P}_\gamma = \left\{ \mathbf{p} : \frac{1}{2} \sum_{\omega \in \Omega} |\mathbf{p}(\omega) - \mathbf{q}(\omega)| \leq \gamma, \quad \sum_{\omega \in \Omega} p(\omega) = 1, \quad \mathbf{p} \geq 0 \right\}.$$