Stochastic programming and robust optimisation

Lecture 3/5

Fabricio Oliveira

Systems Analysis Laboratory
Department of Mathematics and Systems Analysis

Aalto University
School of Science

UFMG, Belo Horizonte, Brazil
27.08.2024

Outline of this lecture

Introduction

Chance constraints

Risk measures

Outline of this lecture

Beyond expected values

Many settings require that a risk profile is imposed:

- Expected values assume a risk neutral stance;
- Risk neutral means that the product probability $\times$ outcome is the single factor under consideration.

Beyond expected values

Many settings require that a risk profile is imposed:

- Expected values assume a risk neutral stance;
- Risk neutral means that the product probability $\times$ outcome is the single factor under consideration.

In many settings, the decision-maker may have other implicit priorities:

- Impose (probabilistic) guarantees on feasibility
  $\Rightarrow$ **chance constraints**;
- Avoid being exposed to the possibility of a high-loss
  $\Rightarrow$ **risk measures**.

Outline of this lecture

Feasibility guarantees

A solution may reveal itself infeasible once the uncertainty unveils

- Particularly critical in settings where infeasibility has a high penalty;
- Appealing in settings in which safety and resilience are requirements.

Feasibility guarantees

A solution may reveal itself infeasible once the uncertainty unveils

- Particularly critical in settings where infeasibility has a high penalty;
- Appealing in settings in which safety and resilience are requirements.

Let us first consider the case without recourse. Our problem is thus

$$\begin{aligned}
\text{min. } & c^\top x \\
\text{s.t.: } & Ax = b \\
& T(\xi)x = h(\xi), \ \forall \xi \in \Xi \\
& x \geq 0.
\end{aligned}$$

Feasibility guarantees

A solution may reveal itself infeasible once the uncertainty unveils

- Particularly critical in settings where infeasibility has a high penalty;
- Appealing in settings in which safety and resilience are requirements.

Let us first consider the case without recourse. Our problem is thus

$$\begin{aligned}
\text{min. } & c^\top x \\
\text{s.t.: } & Ax = b \\
& T(\xi)x = h(\xi), \ \forall \xi \in \Xi \\
& x \geq 0.
\end{aligned}$$

Notice the static nature of the problem

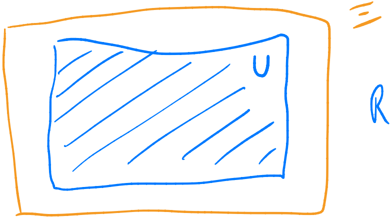
- once a decision x is made, we observe the realisation of the uncertainty ξ;
- no correction decision is allowed.

Feasibility guarantees

There are two main paradigms on how to model feasibility requirements for the constraint $T(\xi)x = h(\xi), \ \forall \xi \in \Xi$:

1. Impose that $T(\xi)x = h(\xi)$ holds for a set of realisations $U \subseteq \Xi$ (robust optimisation)
2. Impose that the probability of $T(\xi)x = h(\xi)$ holding is at least a given threshold (chance constraints)

As one may suspect, the two approaches are interrelated.



Robust // Chance const.

$\mathbb{P}(T(\xi)x = h(\xi)) \geq \alpha$

Feasibility guarantees

There are two main paradigms on how to model feasibility requirements for the constraint $T(\xi)x = h(\xi), \ \forall \xi \in \Xi$:

1. Impose that $T(\xi)x = h(\xi)$ holds for a set of realisations $U \subseteq \Xi$ (robust optimisation)
2. Impose that the probability of $T(\xi)x = h(\xi)$ holding is at least a given threshold (chance constraints)

As one may suspect, the two approaches are interrelated.

Imposing a chance constraint means that a solution is deemed acceptable if, for a given confidence level α, we have that

$$\mathbb{P}(T(\xi)x = h(\xi)) \geq \alpha.$$

Types of chance constraints

Let $T(\xi)$ be a $m_2 \times n_1$ matrix, where $T(\xi)_i$ represents its i^{th}-row, and $h(\xi)$ a m_2 vector with components $h(\xi)_i$.

There are two types of chance constraints:

1. **Individual chance constraints** (ICC):

$$p_i(x) := \mathbb{P}((T(\xi)_i)^\top x = h(\xi)_i) \geq \alpha_i, \ \forall i \in [m_2],$$

Types of chance constraints

Let $T(\xi)$ be a $m_2 \times n_1$ matrix, where $T(\xi)_i$ represents its i^{th}-row, and $h(\xi)$ a m_2 vector with components $h(\xi)_i$.

There are two types of chance constraints:

1. **Individual chance constraints** (ICC):

$$p_i(x) := \mathbb{P}((T(\xi)_i)^\top x = h(\xi)_i) \geq \alpha_i, \ \forall i \in [m_2],$$

JCC

2. **Joint chance constraints** (JCC):

$$p(x) := \mathbb{P}((T(\xi)_i)^\top x = h(\xi)_i, \ \forall i \in [m_2]) \geq \alpha$$

Types of chance constraints

Let $T(\xi)$ be a $m_2 \times n_1$ matrix, where $T(\xi)_i$ represents its i^{th}-row, and $h(\xi)$ a m_2 vector with components $h(\xi)_i$.

There are two types of chance constraints:

1. **Individual chance constraints** (ICC):

$$p_i(x) := \mathbb{P}((T(\xi)_i)^\top x = h(\xi)_i) \geq \alpha_i, \ \forall i \in [m_2],$$

2. **Joint chance constraints** (JCC):

$$p(x) := \mathbb{P}((T(\xi)_i)^\top x = h(\xi)_i, \ \forall i \in [m_2]) \geq \alpha$$

JCCs are typically more challenging from a tractability standpoint

- ICCs can be used to approximate a JCC;
- **Bonferroni inequality**: for a given x, if $p_i(x) > \alpha_i$, $\forall i \in [m_2]$, where $\alpha_i = 1 - \frac{(1-\alpha)}{m_2}$, then $p(x) \geq \alpha$.

Tractability of chance constraints

Tractability issues stem from the properties of the feasible sets generated by chance constraints. Let

$$C(\alpha_1, \dots, \alpha_{m_2}) := \bigcap_{i \in [m_2]} = C_i(\alpha_i), \text{ where}$$

$$C_i(\alpha_i) = \{x \in \mathbb{R}^n : p_i(x) \geq \alpha_i\}.$$

- No general result that guarantees the convexity of $C_i(\alpha_i)$;
- Particular cases do exist for important distributions which lead to the convexity of $C(\alpha_1, \dots, \alpha_{m_2})$.

Tractability of chance constraints

Tractability issues stem from the properties of the feasible sets generated by chance constraints. Let

$$C(\alpha_1, \dots, \alpha_{m_2}) := \bigcap_{i \in [m_2]} = C_i(\alpha_i), \text{ where}$$

$$C_i(\alpha_i) = \{x \in \mathbb{R}^n : p_i(x) \geq \alpha_i\}.$$

- No general result that guarantees the convexity of $C_i(\alpha_i)$;
- Particular cases do exist for important distributions which lead to the convexity of $C(\alpha_1, \dots, \alpha_{m_2})$.

For example, assume that $T(\xi) = T, \forall \xi \in \Xi$, and that $h(\xi) = \xi$. For the univariate case, we have that

$$C(\alpha) = \left\{x \in \mathbb{R}^{n_1} : Tx \geq F^{-1}(\alpha)\right\}.$$

Tractability of chance constraints

One general result that is known and can be informative in the multivariate case is the following:

Theorem 1

Let $T(\xi) = T$, $\forall \xi \in \Xi$, and $h(\xi) = \xi$, where $\xi \in \mathbb{R}^{m_2}$ is a random vector with density function f. If $\log(f)$ is concave (assuming $\log(0) = -\infty$), then $C(\alpha)$ is closed and convex for all $\alpha \in [0, 1]$.

Tractability of chance constraints

One general result that is known and can be informative in the multivariate case is the following:

Theorem 1

Let $T(\xi) = T$, $\forall \xi \in \Xi$, and $h(\xi) = \xi$, where $\xi \in \mathbb{R}^{m_2}$ is a random vector with density function f. If $\log(f)$ is concave (assuming $\log(0) = -\infty$), then $C(\alpha)$ is closed and convex for all $\alpha \in [0, 1]$.

- **Main case:** $\xi \sim \text{Normal}(\mu, \Sigma)$ with vector mean μ and covariance matrix Σ;
- Uniform case also “trivial” to show that holds;
- For a list of “other distributions”: [Nemirovski and Shapiro, 2007]

Tractability of chance constraints

Another important known case is this: $T(\xi)$ is a $1 \times n$ random vector and $h(\xi) = h$, $\forall \xi \in \Xi$.

Theorem 2

Assume that $T(\xi) = \xi = (\xi_i)_i = 1^{n_1}$ *is the only random parameter, where* $\xi \sim$ *Normal*(μ, Σ) *with* $\mu = (\mu_i)_{i=1}^{n_1}$ *a vector of means and* Σ *the covariance matrix. Then*

$$C(\alpha) = \left\{ x \in \mathbb{R}^{n_1} : \mu^\top x \geq h + \Phi^{-1}(\alpha)\sqrt{x^\top \Sigma x} \right\}$$

Tractability of chance constraints

Proof.

The random variable $\xi^\top x$ is a multivariate normal with mean $\mu^\top x$ and variance $x^\top \Sigma x$. Letting Z follow a standard normal, we have that

$$\mathbb{P}(\xi^\top x \geq h) \geq \alpha \Leftrightarrow \mathbb{P}\left(\frac{\xi^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}} \geq \frac{h^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha$$

$\square$

Tractability of chance constraints

Proof.

The random variable $\xi^\top x$ is a multivariate normal with mean $\mu^\top x$ and variance $x^\top \Sigma x$. Letting Z follow a standard normal, we have that

$$\mathbb{P}(\xi^\top x \geq h) \geq \alpha \Leftrightarrow \mathbb{P}\left(\frac{\xi^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}} \geq \frac{h^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha$$

$$\Leftrightarrow 1 - \mathbb{P}\left(Z \leq \frac{h - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha$$

$\square$

Tractability of chance constraints

Proof.

The random variable $\xi^\top x$ is a multivariate normal with mean $\mu^\top x$ and variance $x^\top \Sigma x$. Letting Z follow a standard normal, we have that

$$
\begin{aligned}
\mathbb{P}(\xi^\top x \geq h) \geq \alpha &\Leftrightarrow \mathbb{P}\left(\frac{\xi^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}} \geq \frac{h^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow 1 - \mathbb{P}\left(Z \leq \frac{h - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow 1 - \Phi\left(\frac{h - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha
\end{aligned}
$$

$\square$

Tractability of chance constraints

Proof.

The random variable $\xi^\top x$ is a multivariate normal with mean $\mu^\top x$ and variance $x^\top \Sigma x$. Letting Z follow a standard normal, we have that

$$\begin{aligned}
\mathbb{P}(\xi^\top x \geq h) \geq \alpha &\Leftrightarrow \mathbb{P}\left(\frac{\xi^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}} \geq \frac{h^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow 1 - \mathbb{P}\left(Z \leq \frac{h - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow 1 - \Phi\left(\frac{h - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow \Phi\left(\frac{\mu^\top x - h}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha
\end{aligned}$$

□

Tractability of chance constraints

Proof.

The random variable $\xi^\top x$ is a multivariate normal with mean $\mu^\top x$ and variance $x^\top \Sigma x$. Letting Z follow a standard normal, we have that

$$
\begin{aligned}
\mathbb{P}(\xi^\top x \geq h) \geq \alpha &\Leftrightarrow \mathbb{P}\left(\frac{\xi^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}} \geq \frac{h^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow 1 - \mathbb{P}\left(Z \leq \frac{h - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow 1 - \Phi\left(\frac{h - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow \Phi\left(\frac{\mu^\top x - h}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow \frac{\mu^\top x - h}{\sqrt{x^\top \Sigma x}} \geq \Phi^{-1}(\alpha) \Leftrightarrow \mu^\top x \geq h + \Phi^{-1}(\alpha)\sqrt{x^\top \Sigma x}
\end{aligned}
$$

$\square$

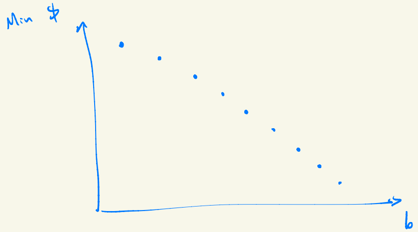
Tractability of chance constraints

Proof.

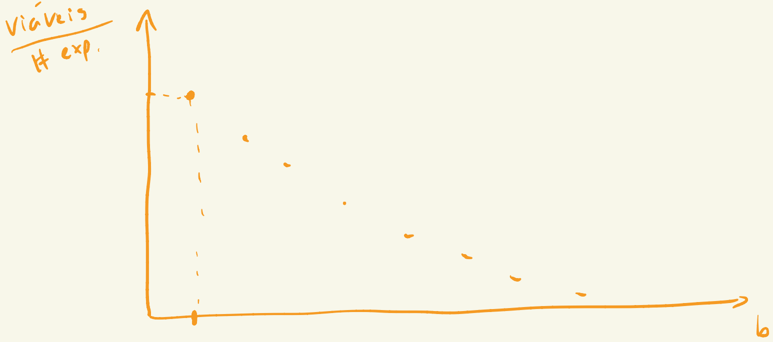
The random variable $\xi^\top x$ is a multivariate normal with mean $\mu^\top x$ and variance $x^\top \Sigma x$. Letting Z follow a standard normal, we have that

$$\begin{aligned}
\mathbb{P}(\xi^\top x \geq h) \geq \alpha &\Leftrightarrow \mathbb{P}\left(\frac{\xi^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}} \geq \frac{h^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow 1 - \mathbb{P}\left(Z \leq \frac{h - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow 1 - \Phi\left(\frac{h - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow \Phi\left(\frac{\mu^\top x - h}{\sqrt{x^\top \Sigma x}}\right) \geq \alpha \\
&\Leftrightarrow \frac{\mu^\top x - h}{\sqrt{x^\top \Sigma x}} \geq \Phi^{-1}(\alpha) \Leftrightarrow \mu^\top x \geq h + \Phi^{-1}(\alpha)\sqrt{x^\top \Sigma x} \qquad \square
\end{aligned}$$

Notice that the constraint is convex if $\Phi^{-1}(\alpha) \geq 0$, i.e., $\alpha \in [1/2, 1]$.



Min $
b



viáveis / # exp.
b

Discretisation of chance constraints

An alternative way of handling chance constraints is using scenarios:

- Allows for general (discrete) distributions;
- Convex problems by construction (for an originally convex problem);
- Allows for recourse decisions;
- Requires binary variables per scenario, which may be an issue computationally.

Bertsimas & Sim (2004)

James Luedtke (Madison)

Discretisation of chance constraints

An alternative way of handling chance constraints is using scenarios:

- Allows for general (discrete) distributions;
- Convex problems by construction (for an originally convex problem);
- Allows for recourse decisions;
- Requires binary variables per scenario, which may be an issue computationally.

Let us consider again our scenario-based deterministic equivalent 2SSP:

$$\begin{aligned}
\text{min. } & c^\top x + \sum_{s \in S} P_s q_s^\top y_s \\
\text{s.t.: } & Ax = b, x \geq 0 \\
& T_s x + W_s y_s = h_s, \ \forall s \in S \\
& y_s \geq 0, \ \forall s \in S.
\end{aligned}$$

Discretisation of chance constraints

Let $v_s \in \{0, 1\}$, $u_s \in \mathbb{R}$, $\forall s \in S$, and M be a sufficiently large (big-M) value.

Discretisation of chance constraints

Let $v_s \in \{0, 1\}$, $u_s \in \mathbb{R}$, $\forall s \in S$, and M be a sufficiently large (big-M) value. Then, we can reformulate our chance-constrained problem as

$$
\begin{aligned}
\text{min. } & c^\top x + \sum_{s \in S} P_s q_s^\top y_s && \text{(1a)}\\
\text{s.t.: } & Ax = b && \text{(1b)}\\
& T_s x + W_s y_s = h_s + u_s, \ \forall s \in S && \text{(1c)}\\
& |u_s| \leq M v_s, \ \forall s \in S && \text{(1d)}\\
& \sum_{s \in S} P_s v_s \leq 1 - \alpha && \text{(1e)}\\
& x \geq 0 && \text{(1f)}\\
& y_s \geq 0, u_s \in \mathbb{R}, v_s \in \{0, 1\} \ \forall s \in S, && \text{(1g)}
\end{aligned}
$$

$|h_s - T_s x + W_s y_s|$

$v_s = \begin{cases} 0, & \text{if (1c) is feasible} \\ 1, & \text{if (1c) is infeasible} \end{cases}$

infeasibility prob.

where α be our feasibility likelihood.

Discretisation of chance constraints

Some final remarks:

- $|u_s| \leq Mv_s$, $\forall s \in S$, can be easily linearised, and is not necessary when the chance-constraints are one-sided (i.e., $\leq$ or $\geq$);

Discretisation of chance constraints

Some final remarks:

- $|u_s| \leq M v_s, \; \forall s \in S$, can be easily linearised, and is not necessary when the chance-constraints are one-sided (i.e., $\leq$ or $\geq$);
- The binary variables v_s indicate which scenarios are infeasible.

Discretisation of chance constraints

Some final remarks:

- $|u_s| \leq Mv_s, \; \forall s \in S$, can be easily linearised, and is not necessary when the chance-constraints are one-sided (i.e., $\leq$ or $\geq$);
- The binary variables v_s indicate which scenarios are infeasible.
- $\sum_{s \in S} P_s v_s$ gives the infeasibilty probability;

Discretisation of chance constraints

Some final remarks:

- $|u_s| \leq Mv_s, \ \forall s \in S$, can be easily linearised, and is not necessary when the chance-constraints are one-sided (i.e., $\leq$ or $\geq$);
- The binary variables v_s indicate which scenarios are infeasible.
- $\sum_{s \in S} P_s v_s$ gives the infeasibilty probability;
- Notice that one binary variable per scenario is required;

Discretisation of chance constraints

Some final remarks:

- $|u_s| \leq Mv_s, \ \forall s \in S$, can be easily linearised, and is not necessary when the chance-constraints are one-sided (i.e., $\leq$ or $\geq$);
- The binary variables v_s indicate which scenarios are infeasible.
- $\sum_{s \in S} P_s v_s$ gives the infeasibilty probability;
- Notice that one binary variable per scenario is required;
- The above can be circumvented using integrated chance constraints.

Discretisation of chance constraints

Some final remarks:

- $|u_s| \leq Mv_s, \; \forall s \in S$, can be easily linearised, and is not necessary when the chance-constraints are one-sided (i.e., $\leq$ or $\geq$);
- The binary variables v_s indicate which scenarios are infeasible.
- $\sum_{s \in S} P_s v_s$ gives the infeasibilty probability;
- Notice that one binary variable per scenario is required;
- The above can be circumvented using integrated chance constraints.
  - Alternative, one can impose limits on the expected infeasibility (variables $u_s, \; \forall s \in S$)

Discretisation of chance constraints

Some final remarks:

- $|u_s| \leq Mv_s, \ \forall s \in S$, can be easily linearised, and is not necessary when the chance-constraints are one-sided (i.e., $\leq$ or $\geq$);
- The binary variables v_s indicate which scenarios are infeasible.
- $\sum_{s \in S} P_s v_s$ gives the infeasibilty probability;
- Notice that one binary variable per scenario is required;
- The above can be circumvented using integrated chance constraints.
  - Alternative, one can impose limits on the expected infeasibility (variables $u_s, \forall s \in S$)
  - This is achieved by replacing (1d) and (1e) with

$$\sum_{s \in S} P_s u_s \leq \beta$$

where β is a limit on the expected amount of infeasibility;

Chance constraints

Outline of this lecture

Introduction

Chance constraints

Risk measures

Measuring risk

Recall that we seek to find a solution x that optimises

$$\min_x \mathbb{E}_\xi \left[F(x, \xi) \right],$$

where:

- $F(x, \xi) = \left\{ c^\top x + Q(x, \xi) : x \in X \right\}$;
- $Q(x, \xi) = \text{min.}_{\ y} \left\{ q(\xi)^\top y : W(\xi) y = h(\xi) - T(\xi) x, y \geq 0 \right\}$;
- $X = \{ x \in \mathbb{R}^n : Ax = b, x \geq 0 \}$.

Measuring risk

Recall that we seek to find a solution x that optimises

$$\min_x \mathbb{E}_\xi \left[F(x,\xi)\right],$$

where:

- $F(x,\xi) = \left\{c^\top x + Q(x,\xi) : x \in X\right\}$;
- $Q(x,\xi) = \text{min.}_{\ y} \left\{q(\xi)^\top y : W(\xi)y = h(\xi) - T(\xi)x, y \geq 0\right\}$;
- $X = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$.

That is, we choose $x^\star = \operatorname{argmin}_x \mathbb{E}_\xi \left[F(x,\xi)\right]$.

- Each $x' \in X$ has an associated a probability distribution $f_x(\xi)$ which maps a cost $F(x',\xi)$ to the probability of scenario ξ;
- Thus, x' is preferred over x'' if $\mathbb{E}_\xi \left[F(x',\xi)\right] < \mathbb{E}_\xi \left[F(x'',\xi)\right]$.

Measuring risk

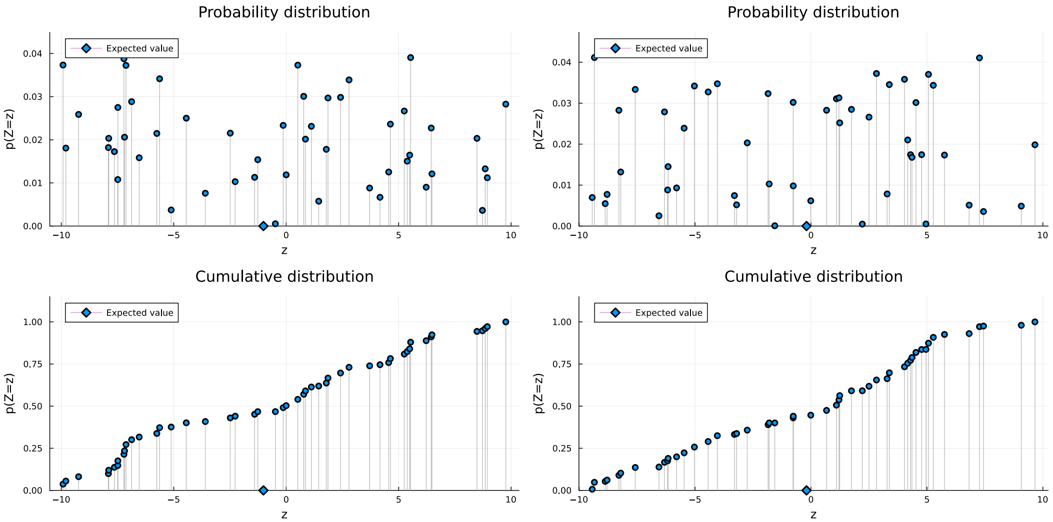
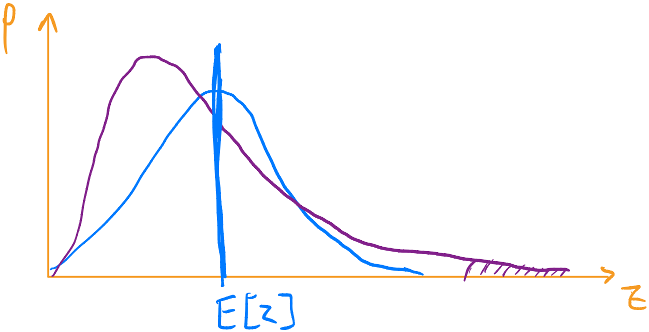


Figure: Comparing two solutions: the solution generating the cost distribution on the left is preferred, as it has a lower expected value

Measuring risk

However, choosing between distributions using their expected values neglects information about the dispersion:

- Higher-order statistical moments are disregarded
- Tails of the cost distribution are often relevant from a decision-making standpoint.



Measuring risk

However, choosing between distributions using their expected values neglects information about the dispersion:

- Higher-order statistical moments are disregarded
- Tails of the cost distribution are often relevant from a decision-making standpoint.

To capture more information about such tails we can define risk measures $r : X \to \mathbb{R}$ such that

- r associates the random variable $F(x, \xi)$ generated by the solution x with a real-valued risk $r_\xi(x)$
- Analogously, x' can be chosen over x'' if $r_\xi\left[F(x', \xi)\right] < r_\xi\left[F(x'', \xi)\right]$.

Trading off risk and expected return

Being two conflicting objectives, risk and return are typically considered under a bi-objective standpoint, e.g., using

1. Weighted terms in the objective function:

$$\min_x \mathbb{E}_\xi \left[F(x,\xi)\right] + \beta r_\xi \left[F(x,\xi)\right],$$

   where $\beta = 0$ represents a risk-neutral stance and risk aversion increases as $\beta \to \infty$;

Trading off risk and expected return

Being two conflicting objectives, risk and return are typically considered under a bi-objective standpoint, e.g., using

1. Weighted terms in the objective function:

$$\min_x \mathbb{E}_\xi \left[F(x,\xi)\right] + \beta r_\xi \left[F(x,\xi)\right],$$

   where $\beta = 0$ represents a risk-neutral stance and risk aversion increases as $\beta \to \infty$;

2. A risk exposition budget δ:

$$\begin{aligned} &\min_x \ \mathbb{E}_\xi \left[F(x,\xi)\right] \\ &\text{s.t.: } r_\xi \left[F(x,\xi)\right] \leq \delta. \end{aligned}$$

Coherent risk measures

[Artzner et al., 1999] provides axiomatic definitions for coherent risk measures:

1. **Translation invariance:** $r_\xi\left[F(x,\xi)+a\right] = r_\xi\left[F(x,\xi)\right] + a$ for $a \in \mathbb{R}$.
2. **Subadditivity:** $r_\xi\left[F(x',\xi)+F(x'',\xi)\right] \leq r_\xi\left[F(x',\xi)\right] + r_\xi\left[F(x'',\xi)\right]$
3. **Positive homogeneity:** $r_\xi\left[F(x,\xi)\times a\right] = r_\xi\left[F(x,\xi)\right] \times a$ for $a \in \mathbb{R}$.
4. **Monotonicity:** if for every ξ, we have that $F(x',\xi) \leq F(x'',\xi)$, then $r_\xi\left[F(x',\xi)\right] \leq r_\xi\left[F(x'',\xi)\right]$

Coherent risk measures

[Artzner et al., 1999] provides axiomatic definitions for coherent risk measures:

1. **Translation invariance:** $r_\xi\left[F(x,\xi)+a\right] = r_\xi\left[F(x,\xi)\right] + a$ for $a \in \mathbb{R}$.
2. **Subadditivity:** $r_\xi\left[F(x',\xi)+F(x'',\xi)\right] \leq r_\xi\left[F(x',\xi)\right] + r_\xi\left[F(x'',\xi)\right]$
3. **Positive homogeneity:** $r_\xi\left[F(x,\xi)\times a\right] = r_\xi\left[F(x,\xi)\right] \times a$ for $a \in \mathbb{R}$.
4. **Monotonicity:** if for every ξ, we have that $F(x',\xi) \leq F(x'',\xi)$, then $r_\xi\left[F(x',\xi)\right] \leq r_\xi\left[F(x'',\xi)\right]$

This has been further developed by [Rockafellar, 2007], who establishes the role of coherence in optimisation problems. A coherent risk measure:

- Preserves convexity;
- Preserves certainty;
- Is insensitive to scaling.

Conditional value-at-risk

The most widespread risk measure in the context of optimisation is the Conditional Value-at-Risk (CVaR)

- A coherent risk measure widely used in other areas;
- Empirical results on its efficacy in production planning: [Alem et al., 2020]

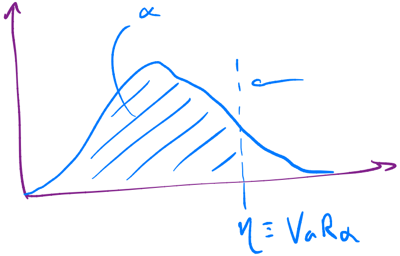
Conditional value-at-risk

The most widespread risk measure in the context of optimisation is the Conditional Value-at-Risk (CVaR)

- A coherent risk measure widely used in other areas;
- Empirical results on its efficacy in production planning: [Alem et al., 2020]

Let X be a random variable and F_X its cumulative distribution function. Then, for a confidence level α, the Value-at-Risk (VaR_α) is defined as

$$VaR_\alpha(X) = \min\{\eta : F_X(\eta) \geq \alpha\}.$$



Conditional value-at-risk

The most widespread risk measure in the context of optimisation is the Conditional Value-at-Risk (CVaR)

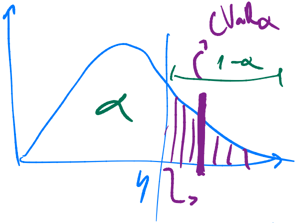
- A coherent risk measure widely used in other areas;
- Empirical results on its efficacy in production planning: [Alem et al., 2020]

Let X be a random variable and F_X its cumulative distribution function. Then, for a confidence level α, the Value-at-Risk (VaR_α) is defined as

$$VaR_\alpha(X) = \min\{\eta : F_X(\eta) \geq \alpha\}.$$

The conditional VaR_α represents the expectation of X in the conditional distribution of its α-upper tail, i.e.,

$$CVaR_\alpha(X) = \min_{\eta \in \mathbb{R}} \left\{ \eta + \frac{\mathbb{E}[X - \eta]^+}{(1 - \alpha)} \right\},$$



where $[\cdot]^+ = \max\{0, \cdot\}$.

Conditional value-at-risk

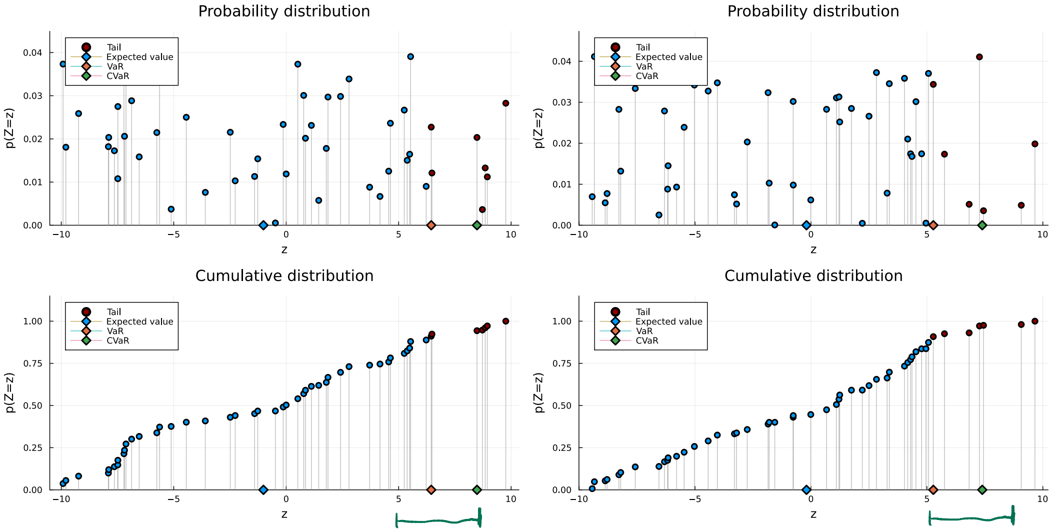


Figure: Comparing two solutions: the solution on the right has better (smaller) $\mathrm{CVaR}_{90\%}$

Conditional value-at-risk in SP models

One appealing feature of CVaR is its convexity:

- Requires discretisation to handle the expected value;
- In the context of optimisation, this means that no additional binary variables are needed;
- This contrasts with VaR (or chance constraints), which need such variables.

Conditional value-at-risk in SP models

One appealing feature of CVaR is its convexity:

- Requires discretisation to handle the expected value;
- In the context of optimisation, this means that no additional binary variables are needed;
- This contrasts with VaR (or chance constraints), which need such variables.

Recall our risk-neutral scenario-based deterministic equivalent 2SSP:

$$\begin{aligned}
\text{min. } & c^\top x + \sum_{s \in S} P_s q_s^\top y_s \\
\text{s.t.: } & Ax = b, x \geq 0 \\
& T_s x + W_s y_s = h_s, \ \forall s \in S \\
& y_s \geq 0, \ \forall s \in S.
\end{aligned}$$

Conditional value-at-risk in SP models

Let us define the following auxiliary terms:

$\min_x \ \mathbb{E}_\xi[F(x,\xi)] + \beta \, r_\xi(x)$

- $\beta \in [0, 1]$: weight for the risk term,
- α: confidence level;
- $\eta \geq 0$: represent the value at risk (VaR);
- $\pi_s \geq 0$, $\forall s \in S$: account for $[X - \eta]^+$. Here $X \equiv c^\top x + q_s^\top y_s$.

"$F(x,\xi)$"

Conditional value-at-risk in SP models

Let us define the following auxiliary terms:

- $\beta \in [0, 1]$: weight for the risk term,
- α: confidence level;
- $\eta \geq 0$: represent the value at risk (VaR);
- $\pi_s \geq 0, \forall s \in S$: account for $[X - \eta]^+$. Here $X \equiv c^\top x + q_s^\top y_s$.

Then, the risk-averse scenario-based deterministic equivalent 2SSP is

$$
\begin{aligned}
\text{min. } & (1-\beta)\left[c^\top x + \sum_{s \in S} P_s q_s^\top y_s\right] + \beta\left[\eta + \frac{\sum_{s \in S} P_s \pi_s}{1-\alpha}\right] \\
\text{s.t.: } & Ax = b, x \geq 0 \\
& T_s x + W_s y_s = h_s, \ \forall s \in S \\
& \pi_s \geq c^\top x + q_s^\top y_s - \eta, \ \forall s \in S \\
& y_s \geq 0, \pi_s \geq 0, \ \forall s \in S \\
& \eta \geq 0.
\end{aligned}
$$

Conditional value-at-risk in SP models

Some final practical remarks:

- Scaling is important. $\beta = 0.5$ is not necessarily a midpoint between risk-neutral and risk-averse solutions.

Conditional value-at-risk in SP models

Some final practical remarks:

- Scaling is important. $\beta = 0.5$ is not necessarily a midpoint between risk-neutral and risk-averse solutions.
- If maximising, pay attention to the sign of the additional terms, as they must change accordingly.

Conditional value-at-risk in SP models

Some final practical remarks:

- Scaling is important. $\beta = 0.5$ is not necessarily a midpoint between risk-neutral and risk-averse solutions.
- If maximising, pay attention to the sign of the additional terms, as they must change accordingly.
- CVaR can also be used in multi-stage problems (see [Shapiro, 2011])

Conditional value-at-risk in SP models

Some final practical remarks:

- Scaling is important. $\beta = 0.5$ is not necessarily a midpoint between risk-neutral and risk-averse solutions.
- If maximising, pay attention to the sign of the additional terms, as they must change accordingly.
- CVaR can also be used in multi-stage problems (see [Shapiro, 2011])
- There are other more recent risk measures for multistage settings ([Dowson et al., 2022]) that:

Conditional value-at-risk in SP models

Some final practical remarks:

- Scaling is important. $\beta = 0.5$ is not necessarily a midpoint between risk-neutral and risk-averse solutions.
- If maximising, pay attention to the sign of the additional terms, as they must change accordingly.
- CVaR can also be used in multi-stage problems (see [Shapiro, 2011])
- There are other more recent risk measures for multistage settings ([Dowson et al., 2022]) that:
  - Behave better (computationally) in dynamic settings

Conditional value-at-risk in SP models

Some final practical remarks:

- Scaling is important. $\beta = 0.5$ is not necessarily a midpoint between risk-neutral and risk-averse solutions.
- If maximising, pay attention to the sign of the additional terms, as they must change accordingly.
- CVaR can also be used in multi-stage problems (see [Shapiro, 2011])
- There are other more recent risk measures for multistage settings ([Dowson et al., 2022]) that:
  - Behave better (computationally) in dynamic settings
  - Serve as proxy to other risk-aversion paradigms (worst-case minimisation or distributional robustness)

References I

Alem, D., Oliveira, F., and Peinado, M. C. R. (2020).
A practical assessment of risk-averse approaches in production lot-sizing problems.
International Journal of Production Research, 58(9):2581–2603.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999).
Coherent measures of risk.
Mathematical finance, 9(3):203–228.

Dowson, O., Morton, D. P., and Pagnoncelli, B. K. (2022).
Incorporating convex risk measures into multistage stochastic programming algorithms.
Annals of Operations Research, pages 1–25.

Nemirovski, A. and Shapiro, A. (2007).
Convex approximations of chance constrained programs.
SIAM Journal on Optimization, 17(4):969–996.

References II

Rockafellar, R. T. (2007).
Coherent approaches to risk in optimization under uncertainty.
In *OR Tools and Applications: Glimpses of Future Technologies*, pages 38–61. Informs.

Shapiro, A. (2011).
Analysis of stochastic dual dynamic programming method.
European Journal of Operational Research, 209(1):63–72.