Distributionally Robust Chance-Constrained Optimization

an overview of optimization under uncertainty

Jake Roth
Advisor: Professor Mihai Anitescu

University of Chicago
Department of Statistics

May 31, 2018

Outline

Motivation and Problem Summary

* Context

Motivation and Problem Summary

* Context
  - Reservoir management
  - Objective: convert water to electricity "optimally"
  - Subject to: environmental constraints, unknown demand, uncertain rainfall
  - Given: historical rainfall and demand data

Motivation and Problem Summary

* Context
  - Reservoir management
  - Objective: convert water to electricity "optimally"
  - Subject to: environmental constraints, unknown demand, uncertain rainfall
  - Given: historical rainfall and demand data

* Goals

Motivation and Problem Summary

* Context
  - Reservoir management
  - Objective: convert water to electricity "optimally"
  - Subject to: environmental constraints, unknown demand, uncertain rainfall
  - Given: historical rainfall and demand data

* Goals
  - Robustness to uncertainty
  - Data-driven solution: make efficient use of independent samples $\xi_1, \ldots, \xi_N$
  - Understand out-of-sample performance: new realizations $\xi^1, \ldots, \xi^M$

Motivation and Problem Summary

* Context
  - Reservoir management
  - Objective: convert water to electricity "optimally"
  - Subject to: environmental constraints, unknown demand, uncertain rainfall
  - Given: historical rainfall and demand data
* Goals
  - Robustness to uncertainty
  - Data-driven solution: make efficient use of independent samples $\xi_1, \ldots, \xi_N$
  - Understand out-of-sample performance: new realizations $\xi^1, \ldots, \xi^M$
* Tools

Motivation and Problem Summary

* Context
  - Reservoir management
  - Objective: convert water to electricity "optimally"
  - Subject to: environmental constraints, unknown demand, uncertain rainfall
  - Given: historical rainfall and demand data
* Goals
  - Robustness to uncertainty
  - Data-driven solution: make efficient use of independent samples $\xi_1, \ldots, \xi_N$
  - Understand out-of-sample performance: new realizations $\xi^1, \ldots, \xi^M$
* Tools
  - Robust optimization (RO)
  - Stochastic optimization (SO)
  - Data-driven optimization (DDO)

Taxonomy of Uncertainty

Taxonomy of Uncertainty

Standard problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \end{aligned} \tag{1}$$

for control variable $x \in \mathbb{R}^n$ and functions $f_i : \mathbb{R}^n \to \mathbb{R}$

Taxonomy of Uncertainty

Standard problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \end{aligned} \tag{1}$$

for control variable $x \in \mathbb{R}^n$ and functions $f_i : \mathbb{R}^n \to \mathbb{R}$

Uncertain parameters

Let $f_1(x) = Ax - b \leq 0$. "Uncertainty" $\approx$ problem data A and b may not be known fully.
Useful distinction: measurement error vs stochastic
Examples, assumptions:

Taxonomy of Uncertainty

Standard problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \dots, m \end{aligned} \qquad (1)$$

for control variable $x \in \mathbb{R}^n$ and functions $f_i : \mathbb{R}^n \to \mathbb{R}$

Uncertain parameters

Let $f_1(x) = Ax - b \le 0$. "Uncertainty" $\approx$ problem data A and b may not be known fully.
Useful distinction: measurement error vs stochastic
Examples, assumptions:

(1) measurement error, but believe $A_{ij} \in [a_{lo}, a_{hi}] = \bar{a}$, $b_i \in [b_{lo}, b_{hi}] = \bar{b}$, $\{\bar{a}, \bar{b}\} := \mathcal{U}$

Taxonomy of Uncertainty

Standard problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \end{aligned} \tag{1}$$

for control variable $x \in \mathbb{R}^n$ and functions $f_i : \mathbb{R}^n \to \mathbb{R}$

Uncertain parameters

Let $f_1(x) = Ax - b \leq 0$. "Uncertainty" $\approx$ problem data A and b may not be known fully.
Useful distinction: measurement error vs stochastic
Examples, assumptions:

(1) measurement error, but believe $A_{ij} \in [a_{lo}, a_{hi}] = \bar{a}, b_i \in [b_{lo}, b_{hi}] = \bar{b}$, $\{\bar{a}, \bar{b}\} := \mathcal{U}$

(2) unobserved process, but must have $A(\omega)x + b(\omega) \leq 0$ for state $\omega \in \Omega \subseteq \mathbb{R}^d$ with dist'n D

Taxonomy of Uncertainty

Standard problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \end{aligned} \tag{1}$$

for control variable $x \in \mathbb{R}^n$ and functions $f_i : \mathbb{R}^n \to \mathbb{R}$

Uncertain parameters

Let $f_1(x) = Ax - b \leq 0$. "Uncertainty" $\approx$ problem data A and b may not be known fully.
Useful distinction: measurement error vs stochastic
Examples, assumptions:

(1) measurement error, but believe $A_{ij} \in [a_{lo}, a_{hi}] = \bar{a}, b_i \in [b_{lo}, b_{hi}] = \bar{b}, \{\bar{a}, \bar{b}\} := \mathcal{U}$

(2) unobserved process, but must have $A(\omega)x + b(\omega) \leq 0$ for state $\omega \in \Omega \subseteq \mathbb{R}^d$ with dist'n D

(3) unobserved process, but believe [4] risk-measure DRO

(a) $(\mathbb{E}[\omega] - \hat{\mu})^T \hat{\Sigma}^{-1} (\mathbb{E}[\omega] - \hat{\mu}) \leq \gamma_{mean}$
(b) $\mathbb{E}\left[(\omega - \hat{\mu})(\omega - \hat{\mu})^T\right] \preceq \gamma_{cov} \hat{\Sigma}$
(c) $\mathbb{E}[\mathbb{1}\{\omega \in \Omega\}] = 1$, Ω closed, convex

Robust Optimization

Robust Optimization

Standard formulation: "optimization for the worst set of parameters"

$$\underset{x}{\text{minimize}} \left\{ \sup_{u \in \mathcal{U}} f_0(x,u) : f_i(x,u) \leq 0, \quad i = 1, \dots, m, \quad \forall u \in \mathcal{U} \right\} \tag{2}$$

for control variable $x \in \mathbb{R}^n$, uncertainty set $\mathcal{U} \ni u$ for parameter element u, and functions $f_i : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$; cardinality of $\mathcal{U}$ may be infinite

Robust Optimization

Standard formulation: "optimization for the worst set of parameters"

$$\underset{x}{\text{minimize}} \left\{ \sup_{u \in \mathcal{U}} f_0(x,u) : f_i(x,u) \leq 0, \quad i = 1, \ldots, m, \quad \forall u \in \mathcal{U} \right\} \tag{2}$$

for control variable $x \in \mathbb{R}^n$, uncertainty set $\mathcal{U} \ni u$ for parameter element u, and functions $f_i : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$; cardinality of $\mathcal{U}$ may be infinite

Robust counterpart

$$\begin{aligned} &\underset{x,t}{\text{minimize}} \{ t : f_0(x,u) \leq t, \ f_i(x,u) \leq 0, \quad i = 1, \ldots, m, \quad \forall u \in \mathcal{U} \} \\ &\underset{x,t}{\text{minimize}} \left\{ t : f_0(x,u) \leq t, \ \sup_{u \in \mathcal{U}} \{ f_i(x,u) \} \leq 0, \quad i = 1, \ldots, m \right\} \end{aligned} \tag{3}$$

Robust Optimization

Standard formulation: "optimization for the worst set of parameters"

$$\underset{x}{\text{minimize}} \left\{ \sup_{u \in \mathcal{U}} f_0(x,u) : f_i(x,u) \leq 0, \quad i = 1, \ldots, m, \quad \forall u \in \mathcal{U} \right\} \tag{2}$$

for control variable $x \in \mathbb{R}^n$, uncertainty set $\mathcal{U} \ni u$ for parameter element u, and functions $f_i : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$; cardinality of $\mathcal{U}$ may be infinite

Robust counterpart

$$\begin{aligned} &\underset{x,t}{\text{minimize}} \{ t : f_0(x,u) \leq t, \ f_i(x,u) \leq 0, \quad i = 1, \ldots, m, \quad \forall u \in \mathcal{U} \} \\ &\underset{x,t}{\text{minimize}} \left\{ t : f_0(x,u) \leq t, \ \sup_{u \in \mathcal{U}} \{ f_i(x,u) \} \leq 0, \quad i = 1, \ldots, m \right\} \end{aligned} \tag{3}$$

Properties

- \+ Safe: Immunize against entire uncertainty set
- \+ Tractable (often): for linear, SOCP, and SDP problems, certain polyhedral sets can preserve the structure of the problem [3]
- \+ One-off interpretable: no reliance on frequentist notion of probability
- – Overly conservative (often): *every* uncertainty realization
- – How to make explicit uncertainty set assumptions?
- – Semi-infinite constraints (but can use duality to convert $\forall$ to $\exists$)

Stochastic Optimization

Stochastic Optimization

General formulation

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & F_0(x,\omega) = \int_\Omega f_0(x,\omega)\ dD(\omega) \\
\text{subject to} \quad & F_i(x,\omega) = \int_\Omega f_i(x,\omega)\ dD(\omega) \leq 0, \quad i = 1,\ldots,m
\end{aligned} \tag{4}
$$

for control variable $x \in \mathbb{R}^n$, uncertainty parameter $\omega \in \mathbb{R}^d$, distribution function D, and constraint functions $f_i : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ [6]

Stochastic Optimization

General formulation

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & F_0(x,\omega) = \int_\Omega f_0(x,\omega)\ dD(\omega) \\ \text{subject to} \quad & F_i(x,\omega) = \int_\Omega f_i(x,\omega)\ dD(\omega) \leq 0, \quad i = 1,\ldots,m \end{aligned} \tag{4}$$

for control variable $x \in \mathbb{R}^n$, uncertainty parameter $\omega \in \mathbb{R}^d$, distribution function D, and constraint functions $f_i : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ [6]

Standard formulation (chance-constraint)

$$\underset{x}{\text{minimize}} \{ f_0(x,\omega) : \ \mathbb{P}[f_i(x,\omega) \leq 0] \geq \alpha, \quad \omega \in \Omega, \quad i = 1,\ldots,m \} \tag{5}$$

Stochastic Optimization

General formulation

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & F_0(x,\omega) = \int_\Omega f_0(x,\omega)\ dD(\omega) \\ \text{subject to} \quad & F_i(x,\omega) = \int_\Omega f_i(x,\omega)\ dD(\omega) \le 0, \quad i = 1,\dots,m \end{aligned} \tag{4}$$

for control variable $x \in \mathbb{R}^n$, uncertainty parameter $\omega \in \mathbb{R}^d$, distribution function D, and constraint functions $f_i : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ [6]

Standard formulation (chance-constraint)

$$\underset{x}{\text{minimize}} \{f_0(x,\omega) : \ \mathbb{P}\left[f_i(x,\omega) \le 0\right] \ge \alpha, \quad \omega \in \Omega, \quad i = 1,\dots,m\} \tag{5}$$

Properties

+ Expressive: CCs operate in the space the decisionmaker can make intuitive sense of
+ Natural: connection to risk measures
− Expensive: quadrature, simulations for integrals?, less-nice distributions?

Example

Example

LP: $x, c, a_i \in \mathbb{R}^n$

$$\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad a_i^T x \le b_i, \quad i = 1, \dots, m \tag{6}$$

Example

LP: $x, c, a_i \in \mathbb{R}^n$

$$\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad a_i^T x \le b_i, \quad i = 1, \dots, m \tag{6}$$

Uncertain LP: RO

Let $\mathcal{U} = \{\{\mathcal{E}_i\}_{i=1}^m\}$, i.e., b, c known, and $\mathcal{E}_i = \{\bar{a}_i + P_i u : \|u\|_2 \le 1\}, \bar{a}_i \in \mathbb{R}^n, P_i \in \mathbb{R}^{n \times n}$

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad a_i^T x \le b_i, \quad \forall a_i \in \mathcal{E}_i, \quad i = 1, \dots, m \\ &\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad \bar{a}_i^T x + \|P_i^T x\|_2 \le b_i, \quad i = 1, \dots, m \end{aligned} \tag{7}$$

Example

LP: $x, c, a_i \in \mathbb{R}^n$

$$\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad a_i^T x \leq b_i, \quad i = 1, \ldots, m \tag{6}$$

Uncertain LP: RO

Let $\mathcal{U} = \{\{\mathcal{E}_i\}_{i=1}^m\}$, i.e., b, c known, and $\mathcal{E}_i = \{\bar{a}_i + P_i u : \|u\|_2 \leq 1\}, \bar{a}_i \in \mathbb{R}^n, P_i \in \mathbb{R}^{n \times n}$

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad a_i^T x \leq b_i, \quad \forall a_i \in \mathcal{E}_i, \quad i = 1, \ldots, m \\ &\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad \bar{a}_i^T x + \|P_i^T x\|_2 \leq b_i, \quad i = 1, \ldots, m \end{aligned} \tag{7}$$

Uncertain LP: SO

Let $a_i \sim N(\bar{a}_i, \Sigma_i)$, i.e., $f_i(x, \omega) = a_i^T x - b_i$ with $D = \Phi$

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad \mathbb{P}\left[a_i^T x \leq b_i\right] \geq \eta, \quad i = 1, \ldots, m \\ &\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad \bar{a}_i^T x + \Phi^{-1}(\eta) \, \|\Sigma_i^{1/2} x\|_2 \leq b_i, \quad i = 1, \ldots, m \end{aligned} \tag{8}$$

Comparison

Comparison

Similarities

Comparison

Similarities

* SO $\rightarrow$ RO: use information about stochastic nature of uncertainty to build $\mathcal{U}$
* RO $\rightarrow$ SO: immunize against all $u \in \mathcal{U}$ to ensure probabilistic coverage in chance-constraint setting

Comparison

Similarities

* SO $\rightarrow$ RO: use information about stochastic nature of uncertainty to build $\mathcal{U}$
* RO $\rightarrow$ SO: immunize against all $u \in \mathcal{U}$ to ensure probabilistic coverage in chance-constraint setting

Differences

Comparison

Similarities

* SO $\rightarrow$ RO: use information about stochastic nature of uncertainty to build $\mathcal{U}$
* RO $\rightarrow$ SO: immunize against all $u \in \mathcal{U}$ to ensure probabilistic coverage in chance-constraint setting

Differences

* "Expressive vs tractable" tradeoff
* RO: probability is not part of the formulation
* SO: immunization only for certain outcomes

Comparison

Similarities

* SO $\rightarrow$ RO: use information about stochastic nature of uncertainty to build $\mathcal{U}$
* RO $\rightarrow$ SO: immunize against all $u \in \mathcal{U}$ to ensure probabilistic coverage in chance-constraint setting

Differences

* "Expressive vs tractable" tradeoff
* RO: probability is not part of the formulation
* SO: immunization only for certain outcomes

Beyond

Comparison

Similarities

* SO $\rightarrow$ RO: use information about stochastic nature of uncertainty to build $\mathcal{U}$
* RO $\rightarrow$ SO: immunize against all $u \in \mathcal{U}$ to ensure probabilistic coverage in chance-constraint setting

Differences

* "Expressive vs tractable" tradeoff
* RO: probability is not part of the formulation
* SO: immunization only for certain outcomes

Beyond

* Beyond RO: "tighten" $\mathcal{U}$ through introducing probabilistic notions
* Beyond SO: generalize by introducing "ambiguity" into chance-constraints
* Constraint form: $\sup_{u \in \mathcal{U}} \{f_i(x, u)\} \leq 0$ vs $\mathbb{P}\left[\{f_i(x, \omega)\} \leq 0\right] \geq \alpha$

Data-driven Optimization

Data-driven Optimization

Typical problem

Data-driven Optimization

Typical problem

* Finite sample $\mathcal{S}$
* RO: estimate $\mathcal{U}$ from $\mathcal{S}$ (build uncertainty sets)
* SO: estimate D from $\mathcal{S}$ (estimate distribution)

Data-driven Optimization

Typical problem

* Finite sample $\mathcal{S}$
* RO: estimate $\mathcal{U}$ from $\mathcal{S}$ (build uncertainty sets)
* SO: estimate D from $\mathcal{S}$ (estimate distribution)

Useful tools

Data-driven Optimization

Typical problem

* Finite sample $\mathcal{S}$
* RO: estimate $\mathcal{U}$ from $\mathcal{S}$ (build uncertainty sets)
* SO: estimate D from $\mathcal{S}$ (estimate distribution)

Useful tools

* Risk-measure literature
* Concentration of measure

Data-driven Optimization

Typical problem

* Finite sample $\mathcal{S}$
* RO: estimate $\mathcal{U}$ from $\mathcal{S}$ (build uncertainty sets)
* SO: estimate D from $\mathcal{S}$ (estimate distribution)

Useful tools

* Risk-measure literature
* Concentration of measure

Interpretation

Data-driven Optimization

Typical problem

* Finite sample $\mathcal{S}$
* RO: estimate $\mathcal{U}$ from $\mathcal{S}$ (build uncertainty sets)
* SO: estimate D from $\mathcal{S}$ (estimate distribution)

Useful tools

* Risk-measure literature
* Concentration of measure

Interpretation

* RO and SO begin to share similar properties in a data-driven context

Risk measures

Risk measures

Definition (Risk measure)

A risk measure $\rho : \mathcal{X} \to \mathbb{R}$, for $\mathcal{X}$ the space of random variables, satisfies

* translation invariance: $\rho(X) = X$ if X is const
* positive homogeneity: $\rho(tX) = t\rho(X)$ for $t > 0$
* normalization: $\rho(0) = 0$
* monotonicity: if $X_1 \leq X_2$ almost surely, then $\rho(X_1) \leq \rho(X_2)$

Risk measures

Definition (Risk measure)

A risk measure $\rho : \mathcal{X} \to \mathbb{R}$, for $\mathcal{X}$ the space of random variables, satisfies

* translation invariance: $\rho(X) = X$ if X is const
* positive homogeneity: $\rho(tX) = t\rho(X)$ for $t > 0$
* normalization: $\rho(0) = 0$
* monotonicity: if $X_1 \leq X_2$ almost surely, then $\rho(X_1) \leq \rho(X_2)$

Definition (VaR, [13])

Let $\alpha \in (0,1)$ be a given confidence level and Z_x be a random variable characterizing the "loss" in a particular system under decision x. Then for cdf F_{Z_x}

$$\mathrm{VaR}_\alpha\left[Z_x\right] := F^{-}_{Z_x}(1-\alpha) = \inf\{t : \mathbb{P}\left[Z_x > t\right] \leq \alpha\}$$

Risk measures

Definition (Risk measure)

A risk measure $\rho : \mathcal{X} \to \mathbb{R}$, for $\mathcal{X}$ the space of random variables, satisfies

- translation invariance: $\rho(X) = X$ if X is const
- positive homogeneity: $\rho(tX) = t\rho(X)$ for $t > 0$
- normalization: $\rho(0) = 0$
- monotonicity: if $X_1 \leq X_2$ almost surely, then $\rho(X_1) \leq \rho(X_2)$

Definition (VaR, [13])

Let $\alpha \in (0, 1)$ be a given confidence level and Z_x be a random variable characterizing the "loss" in a particular system under decision x. Then for cdf F_{Z_x}

$$\mathrm{VaR}_\alpha [Z_x] := F^{-}_{Z_x}(1 - \alpha) = \inf\{t : \mathbb{P}[Z_x > t] \leq \alpha\}$$

Definition (CVaR, [13])

Under the same scenario as VaR, define ($\overset{*}{=}$ for smooth cdf F_{Z_x})

$$\mathrm{CVaR}_\alpha [Z_x] := \inf_{t \in \mathbb{R}} \{t + \alpha^{-1}\mathbb{E}[[Z_x - t]_+]\} \overset{*}{=} \alpha^{-1} \int_{1-\alpha}^{1} \mathrm{VaR}_{1-s} [Z_x] \, ds$$

Coherent risk measures

Coherent risk measures

Definition (Coherent risk measure)

If the problem outcome is convex with respect to the decision, i.e., $f(x)$ convex in x, then a risk measure is called "coherent" if $\rho(f(x))$ is convex in x [12]. Coherent risk measures satisfy the additional property

* subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$

With positive homogeneity & $\lambda \in [0, 1]$, this gives: $\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2)$

Coherent risk measures

Definition (Coherent risk measure)

If the problem outcome is convex with respect to the decision, i.e., $f(x)$ convex in x, then a risk measure is called "coherent" if $\rho(f(x))$ is convex in x [12]. Coherent risk measures satisfy the additional property

* subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$

With positive homogeneity & $\lambda \in [0, 1]$, this gives: $\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2)$

Theorem (Representation of coherent risk measure, [3])

A risk measure ρ is coherent if and only if there exists a family of probability measures $\mathcal{Q}$ such that

$$\rho(X) = \sup_{q \in \mathcal{Q}} \mathbb{E}_q[X]$$

for random variables X in the space of almost surely bounded random variables.

Coherent risk measures

Definition (Coherent risk measure)

If the problem outcome is convex with respect to the decision, i.e., $f(x)$ convex in x, then a risk measure is called "coherent" if $\rho(f(x))$ is convex in x [12]. Coherent risk measures satisfy the additional property

* subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$

With positive homogeneity & $\lambda \in [0, 1]$, this gives: $\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2)$

Theorem (Representation of coherent risk measure, [3])

A risk measure ρ is coherent if and only if there exists a family of probability measures $\mathcal{Q}$ such that

$$\rho(X) = \sup_{q \in \mathcal{Q}} \mathbb{E}_q[X]$$

for random variables X in the space of almost surely bounded random variables.

CVaR properties

Coherent risk measures

Definition (Coherent risk measure)

If the problem outcome is convex with respect to the decision, i.e., $f(x)$ convex in x, then a risk measure is called "coherent" if $\rho(f(x))$ is convex in x [12]. Coherent risk measures satisfy the additional property

* subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$

With positive homogeneity & $\lambda \in [0,1]$, this gives: $\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2)$

Theorem (Representation of coherent risk measure, [3])

A risk measure ρ is coherent if and only if there exists a family of probability measures $\mathcal{Q}$ such that

$$\rho(X) = \sup_{q \in \mathcal{Q}} \mathbb{E}_q[X]$$

for random variables X in the space of almost surely bounded random variables.

CVaR properties

* Convex (linearity of expectation, convexity of $[x-c]_+$) and hence *coherent*

Coherent risk measures

Definition (Coherent risk measure)

If the problem outcome is convex with respect to the decision, i.e., $f(x)$ convex in x, then a risk measure is called "coherent" if $\rho(f(x))$ is convex in x [12]. Coherent risk measures satisfy the additional property

* subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$

With positive homogeneity & $\lambda \in [0, 1]$, this gives: $\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2)$

Theorem (Representation of coherent risk measure, [3])

A risk measure ρ is coherent if and only if there exists a family of probability measures $\mathcal{Q}$ such that

$$\rho(X) = \sup_{q \in \mathcal{Q}} \mathbb{E}_q[X]$$

for random variables X in the space of almost surely bounded random variables.

CVaR properties

* Convex (linearity of expectation, convexity of $[x - c]_+$) and hence *coherent*
* CVaR $\geq$ VaR (more extreme)

Coherent risk measures

Definition (Coherent risk measure)

If the problem outcome is convex with respect to the decision, i.e., $f(x)$ convex in x, then a risk measure is called "coherent" if $\rho(f(x))$ is convex in x [12]. Coherent risk measures satisfy the additional property

* subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$

With positive homogeneity & $\lambda \in [0,1]$, this gives: $\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2)$

Theorem (Representation of coherent risk measure, [3])

A risk measure ρ is coherent if and only if there exists a family of probability measures $\mathcal{Q}$ such that

$$\rho(X) = \sup_{q \in \mathcal{Q}} \mathbb{E}_q[X]$$

for random variables X in the space of almost surely bounded random variables.

CVaR properties

* Convex (linearity of expectation, convexity of $[x-c]_+$) and hence *coherent*
* CVaR $\geq$ VaR (more extreme)
* CVaR is a weighted average of VaR and conditional expectation of losses exceeding VaR; NOT "robust"

Risk Measures for DDO

Risk Measures for DDO

Hypothetical: portfolio optimization

* *Goal*: For decision weights $x \in \mathbb{R}^n$ and RV returns r, ensure that wealth $x^T r \geq \eta$
* *Given*: $[r_1, \ldots, r_m] = R \in \mathbb{R}^{n \times m}$ historical returns

Risk Measures for DDO

Hypothetical: portfolio optimization

* *Goal*: For decision weights $x \in \mathbb{R}^n$ and RV returns r, ensure that wealth $x^T r \geq \eta$
* *Given*: $[r_1, \ldots, r_m] = R \in \mathbb{R}^{n \times m}$ historical returns

RO-inspired models [1]

* *Scenarios (implicit CVaR)*: define $\mathcal{Q} = \text{conv}\{q_1, \ldots, q_l\}$ over "scenarios" $q_1, q_2, \ldots, q_l$ for $q_i \in \Delta^n$ simplex and build $\mathcal{U} = \text{conv}\{Rq : q \in \mathcal{Q}\}$ so $\mathcal{Q}$ generates a coherent risk measure with sup over $\mathcal{Q}$
* *Explicit CVaR*: CVaR defines $\{\mathcal{Q} = q \in \Delta^n : q_i \leq p_i/\alpha\}$ for $p_i = 1/n$ and $\alpha = j/n$, $j \in \mathbb{Z}_+$

Risk Measures for DDO

Hypothetical: portfolio optimization

* *Goal*: For decision weights $x \in \mathbb{R}^n$ and RV returns r, ensure that wealth $x^T r \geq \eta$
* *Given*: $[r_1, \ldots, r_m] = R \in \mathbb{R}^{n \times m}$ historical returns

RO-inspired models [1]

* *Scenarios (implicit CVaR)*: define $\mathcal{Q} = \text{conv}\{q_1, \ldots, q_l\}$ over "scenarios" $q_1, q_2, \ldots, q_l$ for $q_i \in \Delta^n$ simplex and build $\mathcal{U} = \text{conv}\{Rq : q \in \mathcal{Q}\}$ so $\mathcal{Q}$ generates a coherent risk measure with sup over $\mathcal{Q}$
* *Explicit CVaR*: CVaR defines $\{\mathcal{Q} = q \in \Delta^n : q_i \leq p_i/\alpha\}$ for $p_i = 1/n$ and $\alpha = j/n$, $j \in \mathbb{Z}_+$

SO-inspired models

* *"Robust CVaR"*: minimization with ambiguity in mean and covariance [4]
* *Ambiguous chance-constraints*: VaR constraints with unknown distribution risk-measure DRO
* *Scenarios*: estimate empirical distribution robustly (e.g., factor models [7])

Concentration of Measure for DDO

Concentration of Measure for DDO

Theorem (Hoeffding, [9])

Let $X_1, \ldots, X_n$ be independent, bounded random variables such that $X_i \in [a_i, b_i]$ $\forall i = 1, \ldots, n$. Then we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \geq \delta\right) \leq \exp\left(\frac{-2n^2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Concentration of Measure for DDO

Theorem (Hoeffding, [9])

Let $X_1, \dots, X_n$ be independent, bounded random variables such that $X_i \in [a_i, b_i]\ \forall i = 1, \dots, n$. Then we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \geq \delta\right) \leq \exp\left(\frac{-2n^2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Application: inference for stochastic optimization

* Probabilistic bound on difference between empirical estimate of CVaR and true CVaR

Outline

Formulation

Formulation

Overview

* Worst case VaR constraint over family of probability distributions
* Distributionally robust stochastic program
* Bounded support assumption to use concentration inequality

Formulation

Overview

* Worst case VaR constraint over family of probability distributions
* Distributionally robust stochastic program
* Bounded support assumption to use concentration inequality

Formulation

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & \sup_{F \in \mathcal{D}} \mathbb{P}_F \left[f(x, \xi) \leq 0 \right] \geq \alpha \\ & \text{supp}(\mathcal{D}) \subseteq [a, b] \end{aligned} \tag{9}$$

* Control variable: $x \in \mathbb{R}^d$
* Randomness: $\xi \in \mathbb{R}^p$
* Constraint function: $f : \mathbb{R}^{d+p} \to \mathbb{R}$ is convex in x
* Distribution family: $f(x, \xi) \sim F$ for $F \in \mathcal{D}$ with bounded support
* Certainty: $\alpha \in (0, 1)$

Tractable Approximation I

Goal: $\mathbb{P}_F\left[f(x,\xi) \leq 0\right]$ cdf may not be convex, so we seek a reformulation (and follow [11])

Bound the step-function

Rewrite VaR as 0/1 penalty for RV $Z_x := f(x,\xi)$

$$\mathrm{VaR}_\alpha\left[Z_x\right] \leq 0 \iff \mathbb{P}[Z_x \leq 0] \geq 1-\alpha \iff \mathbb{P}[Z_x > 0] \leq \alpha \iff \mathbb{E}[\mathbb{1}\{Z_x > 0\}] \leq \alpha$$

And bound with convex $\psi : \mathbb{R} \to \mathbb{R}$ such that $\psi(tz) \geq \mathbb{1}\{tz > 0\}$ and $t > 0$

Optimize bound

Replace $t = t^{-1}$

$$\mathbb{E}[\psi(t^{-1}Z_x)] \geq \mathbb{E}[\mathbb{1}\{Z_x > 0\}]\,\forall t > 0 \implies \inf_{t>0}\left\{\mathbb{E}[\psi(t^{-1}Z_x)]\right\} \geq \mathbb{E}[\mathbb{1}\{Z_x > 0\}].$$

and note that $\psi(z) = [1+\gamma z]_+$ for $\gamma > 0$ is smallest for functions such that $\psi(0) = 1$

Tractable Approximation II

Ensure convexity

Write as perspective function $(x, t) \mapsto t\psi(x/t)$ by multiplying by t

$$\inf_{t>0} \left\{ t\,\mathbb{E}[\psi(t^{-1}Z_x)] \right\} \leq \alpha t \implies \mathbb{E}[\mathbb{1}\{Z_x > 0\}] \leq \alpha.$$

Rearrange as CVaR

Rearranging the inequality on the left, substituting $\psi(z)$, replacing $t' = -t$, and rescaling by α, we have

$$\begin{aligned}
\inf_{t>0} \left\{ t\,\mathbb{E}[\psi(t^{-1}Z_x)] - \alpha t \right\} &= \inf_{t>0} \left\{ t\,\mathbb{E}[[1 + t^{-1}Z_x)]_+] - \alpha t \right\} \\
&= \inf_{t>0} \{ \mathbb{E}[[t + Z_x]_+] - \alpha t \} \\
&= \inf_{t'<0} \left\{ \mathbb{E}[[Z_x - t']_+] + \alpha t' \right\} = \inf_{t' \in \mathbb{R}} \left\{ \alpha^{-1}\mathbb{E}[[Z_x - t']_+] + t' \right\} \\
&= \mathrm{CVaR}_\alpha\,[Z_x]
\end{aligned}$$

Tractable Approximation III

Estimate generating function bound

Sample average approximation of ψ expectation (called *generating function*)

$$T = \mathbb{E}_F\left[[f(x,\xi) + t]_+\right]$$

and an empirical estimate

$$\hat{T} = \frac{1}{N}\sum_{i=1}^{N}[f(x,\xi_i) + t]_+$$

Bound out-of-sample performance

Using Hoeffding theorem 6, bound probability of "bad" set Ξ_1

$$\Xi_0 := \{\xi \in \Xi : T(\xi) - \hat{T}(\xi) \leq \delta\}$$

$$\Xi_1 := \Omega \backslash \Xi_0 = \{\xi \in \Xi : T(\xi) - \hat{T}(\xi) > \delta\}$$

$$\mathbb{P}(\Xi_1) \leq \exp\left(\frac{-2N\delta^2}{\Gamma^2}\right) \iff 1 - \mathbb{P}(\Xi_1) = \mathbb{P}(\Xi_0) \geq 1 - \exp\left(\frac{-2N\delta^2}{\Gamma^2}\right)$$

where Γ is support bound

Tractable Approximation IV

Summary

$$T \leq \hat{T} + \delta \leq t(1-\alpha)$$

$$\implies \inf_{t>0} \left[\frac{\mathbb{E}_F \left[f(x,\xi) + t \right]_+}{t} \right] \leq \frac{\frac{1}{N}\sum_{i=1}^{N} [f(x,\xi_i) + t]_+ + \delta}{t} \leq 1-\alpha$$

$$\implies \mathbb{P}_F(f(x,\xi) \geq 0) \leq 1-\alpha$$

Outline

Overview

Overview

Problem

Bi-objective problem:

$$\underset{x\in\mathbb{R}^n_+}{\text{maximize}} \quad \mathbb{E}\left[\xi^T x\right] = \mu^T x \quad \& \quad \underset{x\in\mathbb{R}^n_+}{\text{minimize}} R(x) \quad \text{subject to} \quad \mathbb{1}^T x = 1 \quad \& \quad x_i \geq 0,\ i = 1, \ldots, n \tag{10}$$

Overview

Problem

Bi-objective problem:

$$\underset{x\in\mathbb{R}^n_+}{\text{maximize}} \quad \mathbb{E}\left[\xi^T x\right] = \mu^T x \quad \& \quad \underset{x\in\mathbb{R}^n_+}{\text{minimize}} R(x) \quad \text{subject to} \quad \mathbb{1}^T x = 1 \quad \& \quad x_i \geq 0,\ i = 1, \ldots, n \tag{10}$$

Methods

* Markowitz: $R(x) := x^T \Sigma x$ for empirical covariance Σ
* CC (VaR): $R(x) := \mathbb{P}(\xi^T x \leq \rho) \leq \epsilon$ for specified return threshold ρ
* DRCC (approximate CVaR): $R(x) := \mathbb{P}_{F\in\mathcal{D}}(\xi^T x \leq \rho) \leq \epsilon$ for specified return threshold ρ, certainty parameter ϵ, and distributional set $\mathcal{D}$

Overview

Problem

Bi-objective problem:

$$\underset{x\in\mathbb{R}^n_+}{\text{maximize}} \quad \mathbb{E}\left[\xi^T x\right] = \mu^T x \quad \& \quad \underset{x\in\mathbb{R}^n_+}{\text{minimize}} R(x) \quad \text{subject to} \quad \mathbb{1}^T x = 1 \quad \& \quad x_i \geq 0,\ i = 1, \ldots, n \tag{10}$$

Methods

* Markowitz: $R(x) := x^T \Sigma x$ for empirical covariance Σ
* CC (VaR): $R(x) := \mathbb{P}(\xi^T x \leq \rho) \leq \epsilon$ for specified return threshold ρ
* DRCC (approximate CVaR): $R(x) := \mathbb{P}_{F\in\mathcal{D}}(\xi^T x \leq \rho) \leq \epsilon$ for specified return threshold ρ, certainty parameter ϵ, and distributional set $\mathcal{D}$

Metrics

* Empirical distribution for simulated returns
* Empirical distribution tail probability

Simulation

Simulation

Generate data from n hypothetical assets over m days according to the following:

Simulation

Generate data from n hypothetical assets over m days according to the following:

* Normal
  - true covariance matrix $\Sigma_0 \sim IW(2n, \sqrt{n}I)$ (inverse-Wishart)
  - true mean $\mu_0 \sim N(0, I)$
  - n-dimensional asset return vector $\xi \sim N(\mu_0, \Sigma_0)$
  - m observations $\xi_1, \ldots, \xi_m \sim N(\mu_0, \Sigma_0)$

Simulation

Generate data from n hypothetical assets over m days according to the following:

* Normal
  - true covariance matrix $\Sigma_0 \sim IW(2n, \sqrt{n}I)$ (inverse-Wishart)
  - true mean $\mu_0 \sim N(0, I)$
  - n-dimensional asset return vector $\xi \sim N(\mu_0, \Sigma_0)$
  - m observations $\xi_1, \ldots, \xi_m \sim N(\mu_0, \Sigma_0)$
* Beta: m observations from each of n assets distributed as $\text{Beta}(1 + n - i, 1 + i)$ for $i = 1, \ldots, n$.

Simulation

Generate data from n hypothetical assets over m days according to the following:

* Normal
  - true covariance matrix $\Sigma_0 \sim IW(2n, \sqrt{n}I)$ (inverse-Wishart)
  - true mean $\mu_0 \sim N(0, I)$
  - n-dimensional asset return vector $\xi \sim N(\mu_0, \Sigma_0)$
  - m observations $\xi_1, \ldots, \xi_m \sim N(\mu_0, \Sigma_0)$
* Beta: m observations from each of n assets distributed as $\text{Beta}(1 + n - i, 1 + i)$ for $i = 1, \ldots, n$.
* Normal Mixture: repeat Normal procedure for 2-5 Gaussians with mixture probabilities drawn from normalized uniform variates

Goal:

Simulation

Generate data from n hypothetical assets over m days according to the following:

* Normal
  - true covariance matrix $\Sigma_0 \sim IW(2n, \sqrt{n}I)$ (inverse-Wishart)
  - true mean $\mu_0 \sim N(0, I)$
  - n-dimensional asset return vector $\xi \sim N(\mu_0, \Sigma_0)$
  - m observations $\xi_1, \dots, \xi_m \sim N(\mu_0, \Sigma_0)$
* Beta: m observations from each of n assets distributed as Beta$(1 + n - i, 1 + i)$ for $i = 1, \dots, n$.
* Normal Mixture: repeat Normal procedure for 2-5 Gaussians with mixture probabilities drawn from normalized uniform variates

Goal: ensure that we achieve in excess of $\rho = 1/3 \times \hat{\mu}$ for $\hat{\mu}$ the unconstrained, expected return for new samples

Results I

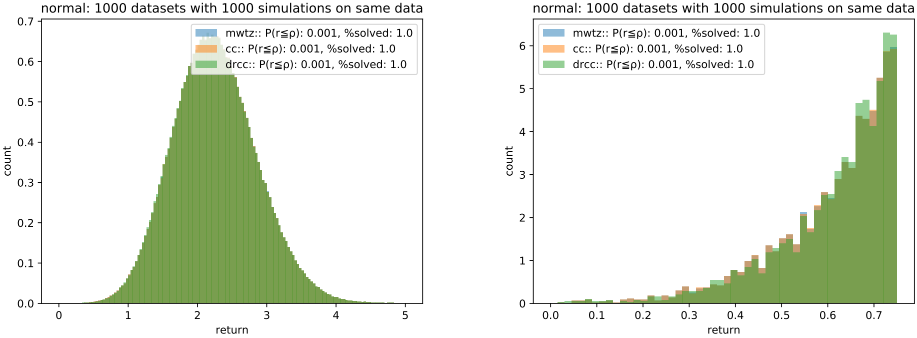


Figure: Normal simulation: we observe nearly identical performance from all three methods

Results II

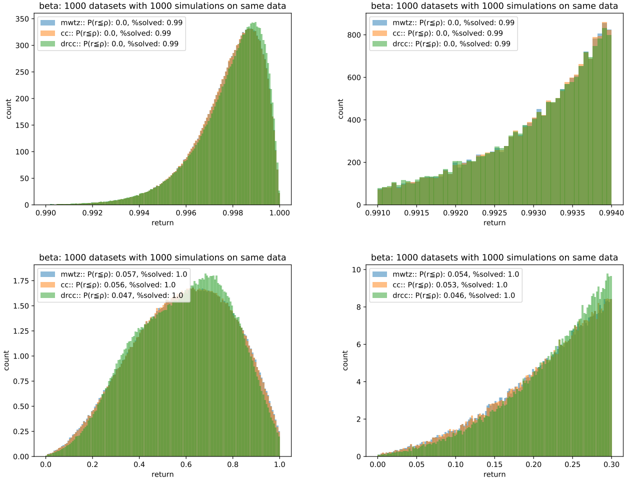
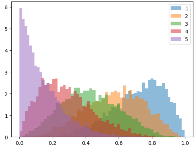
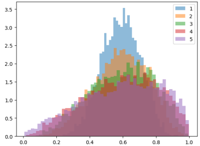


Figure: Beta simulation

- *Top simulations*: $\xi_i \sim \text{Beta}(1+n-i, 1+i)$ for $i = 1, \ldots, n$



- Clear which asset to choose; DRCC shows strong upside
- *Bottom simulations*: $\xi_i \sim \text{Beta}(10/i, 7/i)$ for $i = 1, \ldots, n$



- Clear how to be conservative; DRCC shows strong downside prevention

Results III

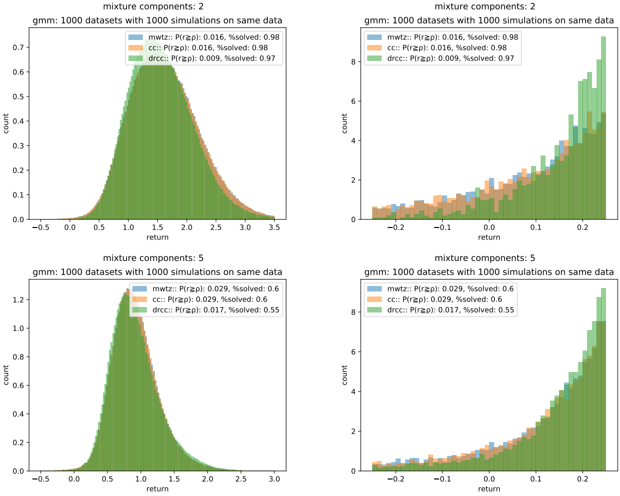


Figure: GMM simulation study, top: 2 mixture components, bottom: 5 mixture components

- ∗ *Top simulations*: DRCC pushes mass out of the tail where $x^T\xi \leq \rho$
- ∗ *Bottom simulations*: DRCC pushes mass out of the tail where $x^T\xi \leq \rho$ and into the better portfolios
- ∗ *Not investigated*: when does DRCC shift "bad" tail mass to "good" tail mass relative to CC and Markowitz?

Overview

Overview

Analytic Problem

Choose simple objective so that we can focus on the constraint

$$\begin{aligned} \underset{x\in[-2,2]}{\text{minimize}} \quad & f_0(x) = x \\ \text{subject to} \quad & \sup_{F\in\mathcal{D}} \mathbb{P}_F\left[\exists u : y(x,u,\xi) := 1 + \xi + x\sin(u) \leq 0\right] \leq \epsilon \\ & u \in [0, 2\pi], \quad \xi \sim \mathcal{D}', \quad \text{supp}(\mathcal{D}') \subseteq [0,1] \end{aligned} \tag{11}$$

Overview

Analytic Problem

Choose simple objective so that we can focus on the constraint

$$\begin{aligned} \underset{x\in[-2,2]}{\text{minimize}} \quad & f_0(x) = x \\ \text{subject to} \quad & \sup_{F\in\mathcal{D}} \mathbb{P}_F\left[\exists u : y(x,u,\xi) := 1 + \xi + x\sin(u) \le 0\right] \le \epsilon \\ & u \in [0, 2\pi], \quad \xi \sim \mathcal{D}', \quad \text{supp}(\mathcal{D}') \subseteq [0,1] \end{aligned} \tag{11}$$

Goals

* Approximate eq. (11) and solve DRCC problem (relaxed, computationally tractable version)
* Attribution and sensitivity analysis for DRCC problem
* Most adverse distribution for analytic problem
* Most adverse distribution for DRCC approximation relative to analytic problem

DRCC Formulation

DRCC Formulation

Solver

$$T := \inf_{t>0} \left[\frac{1}{t} \mathbb{E}_F \left[[-(1 + \xi + \min_{u \in [0,2\pi]} \{x \sin(u)\}) + t]_+ \right] \right]$$

$$\hat{T} := \frac{1}{t} \left[\frac{1}{N} \sum_{i=1}^{N} [-(1 + \xi_i + \min_{u \in [0,2\pi]} \{x \sin(u)\}) + t]_+ + \delta \right]$$

$$\hat{\hat{T}} := \frac{1}{t} \left[\frac{1}{N} \sum_{i=1}^{N} [-(1 + \xi_i + \min_{j=1,\dots,m} \{x \sin(u_j)\}) + t]_+ + \delta \right]$$

$$T \le \hat{T} \quad \text{w.p. } q \ge 1 - \exp\left(\frac{-2N\delta^2}{\Gamma^2} \right)$$

Using $\hat{\hat{T}} \ge \hat{T}$ gives same probabilistic relationship for approximate model

Attribution

Uniform distribution

* Analytic VaR solution: $F^{-1}\left(\mathbb{P}_F\left[\xi < |x| - 1\right]\right) \leq F^{-1}(\epsilon) \implies x^* = -(1 + F^{-1}(\epsilon))$
* Analytic CVaR solution: $x^* = -(1 + \epsilon/2)$

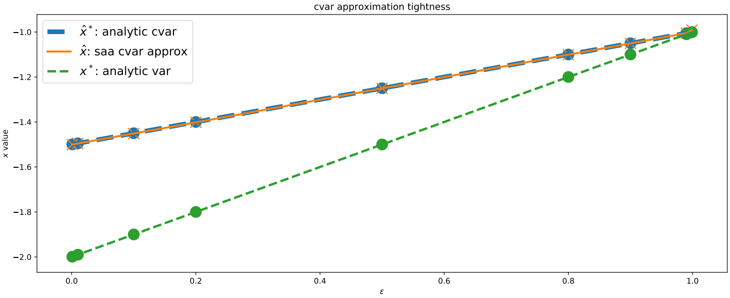


Figure: Effect of analytic vs approximate CVaR. The blue line shows the solution $\hat{x}^*$ of the CVaR problem using analytic CVaR; the orange line shows the solution $\hat{x}$ of the CVaR problem using a sample-average CVaR approximation; the green line shows the solution x^* of the VaR problem using analytic VaR. We observe that the CVaR/VaR relaxation is loose and that the sample-average approximation is reasonably tight.

Attribution

Uniform distribution

* Analytic VaR solution: $F^{-1}(\mathbb{P}_F[\xi < |x| - 1]) \leq F^{-1}(\epsilon) \implies x^* = -(1 + F^{-1}(\epsilon))$
* Analytic CVaR solution: $x^* = -(1 + \epsilon/2)$

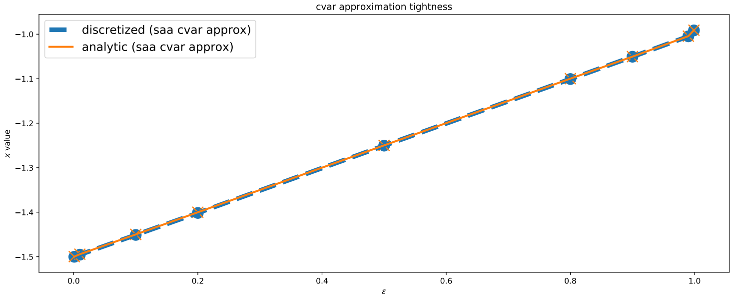


Figure: Effect of analytic vs discretization. The blue line shows the solution of the CVaR problem using a sample-average CVaR approximation and the discretized solution of $\min_u x \sin(u)$; the orange line shows the solution of the CVaR problem using a sample-average CVaR approximation and the analytic solution of $\min_u x \sin(u)$. We observe that the discretization approximation is reasonably tight.

Adverse Distribution (Analytic VaR Problem)

Adverse Distribution (Analytic VaR Problem)

Problem

Solve $\sup_{F \in \mathcal{D}} \{ \mathbb{P}_F [\xi < |x| - 1] \} \leq \epsilon$ for F^*

Adverse Distribution (Analytic VaR Problem)

Problem

Solve $\sup_{F\in\mathcal{D}}\{\mathbb{P}_F[\xi < |x|-1]\} \le \epsilon$ for F^*

Approach: solve VaR via cdf

Characterize F by its left quantile function $F^-(t) = \inf\{z : F(z) \ge t\}$ so that

$$\sup_{F\in\mathcal{D}}\{\mathbb{P}_F[\xi < |x|-1]\} = \sup_{F\in\mathcal{D}}\{F(|x|-1)\} \le \epsilon \tag{12}$$

$$\implies F(|x|-1) \le \epsilon \quad \forall F\in\mathcal{D} \implies |x| \le F^-(\epsilon)+1 \quad \forall F\in\mathcal{D} \implies |x| \le \inf_{F\in\mathcal{D}}\{F^-(\epsilon)\}+1$$

Adverse Distribution (Approximate CVaR Problem)

Problem

Restrict to $\xi \sim \text{Bern}(p)$ and find worst p so that solution to

$$
\begin{aligned}
\underset{x,t}{\text{minimize}} \quad & x \\
\text{subject to} \quad & \frac{1}{N}\sum_{i=1}^{n}[-(1+\xi-|x|)+t]_+ + \delta - t\epsilon \leq 0 \\
& -t \leq 0, \quad |x| \leq 2
\end{aligned}
\tag{13}
$$

is (1) overly-conservative or (2) overly-aggressive relative to analytic solution

Approach: p vs quantile

Given $x^* = -(1+F^{-1}(\epsilon)$, where $F^{-1}(e) = \mathbb{1}\{e > 1-p\}$ find p so that at $p \approx \epsilon$ we have

- $\hat{x} < x^*$: choosing $p > 1-\epsilon \implies x^* = -2$ but approximation is conservative $\hat{x} = 1 - \delta/(1-\hat{p}_1)$ (can solve $(c_1 = 0)$)
- $\hat{x} > x^*$: choosing $p \leq 1-\epsilon$ restricts the analytic solution to -1, but choosing $p \lessapprox 1-\epsilon$ may generate datasets where $\hat{p}_1 > 1-\epsilon$ giving $\hat{x} < -1$ (can solve $(c_1) = (c_3)$)

Adverse Distribution (Approximate CVaR Problem)

Problem

Restrict to $\xi \sim \text{Bern}(p)$ and find worst p so that solution to

$$\begin{aligned}
\underset{x,t}{\text{minimize}} \quad & x \\
\text{subject to} \quad & \hat{p}_1(t+|x|-2) + \hat{p}_0(t+|x|-1) + \delta - \epsilon t \le 0 && (c_1) \\
& \hat{p}_1(t+|x|-2) + \delta - \epsilon t \le 0 && (c_2) \\
& \hat{p}_0(t+|x|-1) + \delta - \epsilon t \le 0 && (c_3) \\
& \delta - \epsilon t \le 0 && (c_4) \\
& -t \le 0, \quad |x| \le 2
\end{aligned} \tag{14}$$

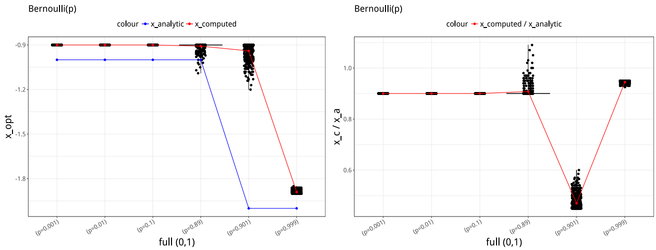
is (1) overly-conservative or (2) overly-aggressive relative to analytic solution

Approach: p vs quantile

Given $x^* = -(1 + F^{-1}(\epsilon)$, where $F^{-1}(e) = \mathbb{1}\{e > 1-p\}$ find p so that at $p \approx \epsilon$ we have

* $\hat{x} < x^*$: choosing $p > 1-\epsilon \implies x^* = -2$ but approximation is conservative $\hat{x} = 1 - \delta/(1-\hat{p}_1)$ (can solve $(c_1 = 0)$)
* $\hat{x} > x^*$: choosing $p \le 1-\epsilon$ restricts the analytic solution to -1, but choosing $p \lessapprox 1-\epsilon$ may generate datasets where $\hat{p}_1 > 1-\epsilon$ giving $\hat{x} < -1$ (can solve $(c_1) = (c_3)$)

Empirical Results



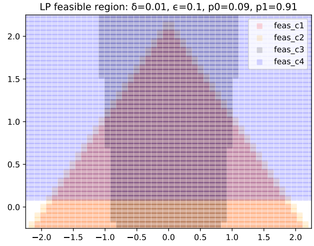


Figure: top: varying p for fixed ϵ, bottom: overly-aggressive approximation

Outline

DRCC Conclusions

Computation

* Tractable for small-medium sized problems
* Require large historical samples to approximate tail expectation
* Robust to outliers or influential samples?

Empirical

* Outperformed standard (limited-assumption) techniques on portfolio problem across distributions
* Bounded support not much of an issue for feasibility (provided enough samples)

Analytic

* CVaR is a useful tool and starting point
* Duality theory
* Combines optimization, statistics, probability

Recent Work

Flavor

* Regularization framework: CVaR as expectation and mean deviations, robust CVaR [7]
* Empirical process theory
  - empirical likelihood confidence intervals related to finding uncertainty sets given by KL-div arguments [5]
  - general conditions under which robust solutions are consistent [5]
* Hypothesis testing for uncertainty/ambiguity sets
  - safe-approximation to ambiguous chance constraints by bounding VaR with a ψ approximation and finding corresponding convex set through duality (epigraph)
  - null hypothesis $F = F_0$ and distributions which pass certain hypothesis tests (e.g., Pearson χ^2, or KL-div tests) at α level define $\mathcal{U}$ [2]
* CVaR / Wasserstein ball
  - set Wasserstein balls around an empirical data-based distribution which allows controllable conservativeness by adjusting the Wasserstein radius [8]
  - Wasserstein ambiguity set centered at empirically estimated distribution [10]

References I

D. Bertsimas and D. Brown, *Constructing uncertainty sets for robust linear optimization*, Operations Research, 57 (2009), pp. 1483–1495.

D. Bertsimas, V. Gupta, and N. Kallus, *Data-driven robust optimization*, Mathematical Programming, (2017).

D. Bertsimas and A. Thiele, *Robust and data-driven optimization: Modern decision-making under uncertainty*, Informs: Tutorials in Operations Research, (2014), pp. 95–122.

E. Delage and Y. Ye, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Operations Research, 58 (2010), pp. 595–612.

J. Duchi, P. Glynn, and H. Namkoong, *Statistics of robust optimization: A generalized empirical likelihood approach*, (2016).

Y. Emoliev and R. J.-B. Wets, *Numerical Techniques for Stochastic Optimization*, Springer-Verlag, 1988.

J.-Y. Gotoh, K. Shinozaki, and A. Takeda, *Robust portfolio techniques for mitigating the fragility of cvar minimization and generalization to coherent risk measures*, Quantitative Finance, 13 (2013), pp. 1621–1635.

References II

- Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, and T. Summers, *Stochastic optimal power flow based on data-driven distributionally robust optimization*, (2017).
- W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association, 58 (1963), pp. 13–30.
- P. Mohajerin Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations*, Mathematical Programming, (2017).
- A. Nemirovski and A. Shaprio, *Convex approximations of chance constraints*, SIAM Journal of Optimization, 17 (2006), pp. 969–996.
- R. T. Rockafeller and J. O. Royset, *Random variables, monotone relations, and convex analysis*, Springer, 148 (2014), p. 297.
- A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on Stochastic Programming*, Society for Industrial and Applied Mathematics and the Mathematical Programming Society, 2009.