Chance Constrained Optimization and their Distributionally Robust Counterpart

using Maximum Mean Discrepancy

Yassine Nemmour[1], Jia-Jie Zhu[2]

[1]Empirical Inference Department
Max-Planck-Institute for Intelligent Systems

[2]Weierstrass Institute for Applied Analysis and Stochastics

Decision-making under uncertainty

$$\min_{x \in \mathcal{X}} c^T x \quad \text{s.t. } f(x,\xi) \leq 0, \quad \xi \in \Xi, \quad \xi \sim P$$

Challenges in Stochastic Optimization

Chance constraints

Decision-making under uncertainty

$$\min_{x \in \mathcal{X}} c^T x \quad \text{s.t. } f(x, \xi) \leq 0, \quad \xi \in \Xi, \quad \xi \sim P$$

- Robust optimization:

$$\min_{x \in \mathcal{X}} c^T x \quad \text{s.t. } f(x, \xi) \leq 0 \quad \forall \xi \in \Xi$$

  $\rightarrow$ Scenario optimization (data-driven)
  Enforce constraint at every sample: $f(x, \xi_i) \leq 0 \quad i = 1, \ldots, N$

Challenges in Stochastic Optimization

Chance constraints

Decision-making under uncertainty

$$\min_{x\in\mathcal{X}} c^T x \quad \text{s.t. } f(x,\xi) \le 0, \quad \xi \in \Xi, \quad \xi \sim P$$

- Robust optimization:

$$\min_{x\in\mathcal{X}} c^T x \quad \text{s.t. } f(x,\xi) \le 0 \quad \forall \xi \in \Xi$$

  $\rightarrow$ Scenario optimization (data-driven)
  Enforce constraint at every sample: $f(x,\xi_i) \le 0 \quad i = 1,\ldots,N$
- Chance constraint to handle very unlikely events:

$$\min_{x\in\mathcal{X}} c^T x \quad \text{s.t. } P(f(x,\xi) \le 0) \ge 1-\alpha \quad \forall \xi \in \Xi$$

Decision-making under uncertainty

$$\min_{x \in \mathcal{X}} c^T x \quad \text{s.t. } f(x, \xi) \leq 0, \quad \xi \in \Xi, \quad \xi \sim P$$

- Robust optimization:

$$\min_{x \in \mathcal{X}} c^T x \quad \text{s.t. } f(x, \xi) \leq 0 \quad \forall \xi \in \Xi$$

  $\rightarrow$ Scenario optimization (data-driven)
  Enforce constraint at every sample: $f(x, \xi_i) \leq 0 \quad i = 1, \ldots, N$
- Chance constraint to handle very unlikely events:

$$\min_{x \in \mathcal{X}} c^T x \quad \text{s.t. } P(f(x, \xi) \leq 0) \geq 1 - \alpha \quad \forall \xi \in \Xi$$

- Distributionally robust chance constraint
  - Assume distributional ambiguity in chance constraint
  - Enforce constraint for worst-case distribution within set of distributions

Convex approximation

Conditional Value-at-Risk

$$P[f(x,\xi) \leq 0] \geq 1-\alpha \quad \Longleftrightarrow \quad \mathrm{VaR}^{P}_{1-\alpha}[f(x,\xi)] \leq 0$$

$$\mathrm{VaR}^{P}_{1-\alpha}[X] = F_X^{-1}(1-\alpha)$$
$$\mathrm{CVaR}^{P}_{1-\alpha}[X] = \mathbb{E}[X | X > \mathrm{VaR}^{P}_{1-\alpha}(X)]$$

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

Convex approximation

Conditional Value-at-Risk

$$P[f(x,\xi) \leq 0] \geq 1-\alpha \quad \Longleftrightarrow \quad \mathsf{VaR}^P_{1-\alpha}[f(x,\xi)] \leq 0$$

$$\mathsf{VaR}^P_{1-\alpha}[X] = F_X^{-1}(1-\alpha)$$
$$\mathsf{CVaR}^P_{1-\alpha}[X] = \mathbb{E}[X | X > \mathsf{VaR}^P_{1-\alpha}(X)]$$

$$\mathsf{VaR}^P_{1-\alpha}[f(x,\xi)] \leq \mathsf{CVaR}^P_{1-\alpha}[f(x,\xi)]$$

Convex approximation

Conditional Value-at-Risk

$$P[f(x,\xi) \le 0] \ge 1-\alpha \quad \Longleftrightarrow \quad \mathrm{VaR}^P_{1-\alpha}[f(x,\xi)] \le 0$$

$$\mathrm{VaR}^P_{1-\alpha}[X] = F_X^{-1}(1-\alpha)$$
$$\mathrm{CVaR}^P_{1-\alpha}[X] = \mathbb{E}[X | X > \mathrm{VaR}^P_{1-\alpha}(X)]$$

$$\mathrm{VaR}^P_{1-\alpha}[f(x,\xi)] \le \mathrm{CVaR}^P_{1-\alpha}[f(x,\xi)]$$

Chance constraint

$$\min_{x \in \mathcal{X}} \quad c^T x$$
$$\text{s.t.} \quad \mathrm{VaR}^P_{1-\alpha}[f(x,\xi)] \le 0$$

Convex approximation

Conditional Value-at-Risk

$$P[f(x,\xi) \le 0] \ge 1-\alpha \quad \Longleftrightarrow \quad \mathrm{VaR}^P_{1-\alpha}[f(x,\xi)] \le 0$$

$$\mathrm{VaR}^P_{1-\alpha}[X] = F_X^{-1}(1-\alpha)$$
$$\mathrm{CVaR}^P_{1-\alpha}[X] = \mathbb{E}[X | X > \mathrm{VaR}^P_{1-\alpha}(X)]$$

$$\mathrm{VaR}^P_{1-\alpha}[f(x,\xi)] \le \mathrm{CVaR}^P_{1-\alpha}[f(x,\xi)]$$

Chance constraint

$$\min_{x\in\mathcal{X}} \quad c^T x$$
$$\text{s.t.} \quad \mathrm{VaR}^P_{1-\alpha}[f(x,\xi)] \le 0$$

CVaR constraint

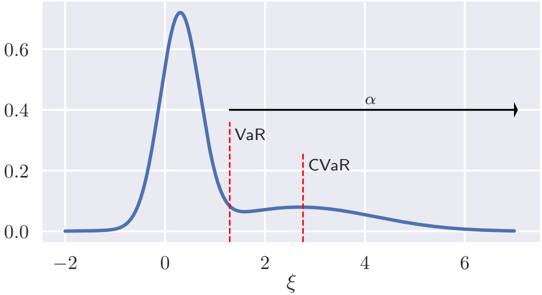
$$\min_{x\in\mathcal{X}} \quad c^T x$$
$$\text{s.t.} \quad \mathrm{CVaR}^P_{1-\alpha}[f(x,\xi)] \le 0$$

Chance constraints

How to solve in practice

$$VaR^P_{1-\alpha} := \inf_{t\in\mathbb{R}} P[f(x,\xi) \leq t] \geq 1-\alpha$$

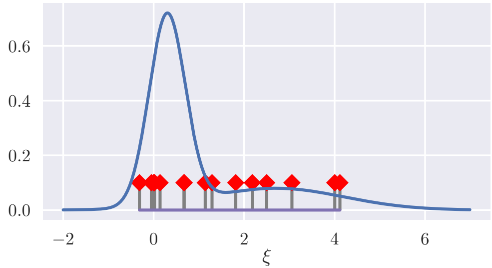
$$CVaR^P_{1-\alpha} := \inf_{t\in\mathbb{R}} \mathbb{E}_P[[f(x,\xi)+t]_+ - t\alpha$$



Distributionally Robust Chance Constraints

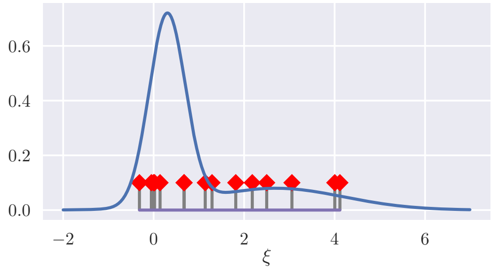
DRO

Distributional ambiguity since we only have samples available



- $\hat{P} = \sum_{i=1}^{N} \frac{1}{N} \delta_{\xi_i}$
- $\hat{P}$ is only a approximation of P_{true}

Distributional ambiguity since we only have samples available



- $\hat{P} = \sum_{i=1}^{N} \frac{1}{N}\delta_{\xi_i}$
- $\hat{P}$ is only a approximation of P_{true}

Ambiguity set: Metric ball around $\hat{P}$ for worst-case distribution

$$\mathcal{P} := \{P : ||P - \hat{P}|| \leq \epsilon\}$$

$$P(f(x,\xi) \leq 0) \geq 1 - \alpha \quad \rightarrow \quad \inf_{P \in \mathcal{P}} P(f(x,\xi) \leq 0) \geq 1 - \alpha$$

- If ϵ large enough $P_{true} \in \mathcal{P}$

Ambiguity set
Definition

$$\mathcal{P} := \left\{ P : ||P - \hat{P}|| \leq \epsilon \right\}$$

Wasserstein distance

$$||P - \hat{P}|| = W_p(P, \hat{P}) = \left(\inf_{\pi \in \Pi(P, \hat{P})} \int ||\xi - \hat{\xi}||^p \pi(d\xi, d\hat{\xi}) \right)^{\frac{1}{p}}$$

- Cross-validation: Very costly for large problems (Chen u. a. 2022; Weijun 2018)
- Empirical likelihood: Limitation w.r.t. constraint function (Blanchet u. a. 2019)

Ambiguity set
Definition

$$\mathcal{P} := \left\{ P : ||P - \hat{P}|| \leq \epsilon \right\}$$

Maximum Mean Discrepancy (MMD)

$$||P - \hat{P}|| = \mathrm{MMD}(P, Q) := ||\mu_P - \mu_Q||_{\mathcal{H}}$$

Kernel Mean Embedding: $\mu_P(\cdot) = \int_{\mathcal{X}} k(x, \cdot) dP(x)$

Ambiguity set construction

$$\mathcal{P} := \{P : \text{MMD}(P, \hat{P}) \leq \epsilon\}$$

with

$$\text{MMD}(P, Q) = \sqrt{\mathbb{E}_P k(x, x') + \mathbb{E}_Q k(y, y') - 2\mathbb{E}_{x\sim P, y\sim Q} k(x, y)}$$

- Closed-form estimator through above definition
- Concentration rates

Ambiguity set construction

$$\mathcal{P} := \{P : \text{MMD}(P, \hat{P}) \leq \epsilon\}$$

with

$$\text{MMD}(P, Q) = \sqrt{\mathbb{E}_P k(x, x') + \mathbb{E}_Q k(y, y') - 2\mathbb{E}_{x \sim P, y \sim Q} k(x, y)}$$

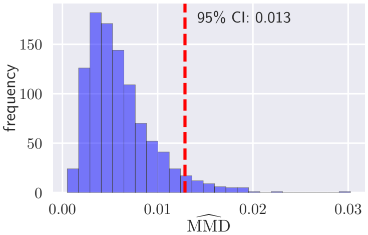
- Closed-form estimator through above definition
- Concentration rates

Bootstrap estimation of MMD

Let $\{\tilde{\xi}_i\}_{i=1}^N$ be a bootstrap sample from $\hat{P}$ with distribution $\tilde{P}$

$$\text{MMD}(\tilde{P}, \hat{P}) \xrightarrow{p} \text{MMD}(P_{True}, \hat{P})$$

$\rightarrow$ Not possible with Wasserstein distances



Distributionally Robust Chance Constraints

Exact reformulation

Primal formulation of feasible set:

$$
\begin{aligned}
Z &:= \{x \in \mathcal{X} : \inf_{P \in \mathcal{P}} P(f(x,\xi) \le 0) \ge 1 - \alpha\} \\
&= \{x \in \mathcal{X} : \sup_{P \in \mathcal{P}} \mathbb{E}_P[\mathbb{1}(f(x,\xi) \ge 0))] \le \alpha\}
\end{aligned}
$$

Distributionally Robust Chance Constraints

Exact reformulation

Primal formulation of feasible set:

$$\begin{aligned} Z &:= \{x \in \mathcal{X} : \inf_{P \in \mathcal{P}} P(f(x,\xi) \leq 0) \geq 1-\alpha\} \\ &= \{x \in \mathcal{X} : \sup_{P \in \mathcal{P}} \mathbb{E}_P[\mathbb{1}(f(x,\xi) \geq 0))] \leq \alpha\} \end{aligned}$$

Assumption: Constraint function $f(x,\xi)$ is semi-continuous

Dual formulation of feasible set (Zhu u. a. 2020):

$$Z := \left\{ x \in \mathcal{X} : \begin{array}{l} g_0 + \frac{1}{N}\sum_{i=0}^{N} g(\xi_i) + \varepsilon ||g||_{\mathcal{H}} \leq \alpha \\ \mathbb{1}(f(x,\xi) > 0) \leq g(\xi) + g_0 \quad \forall \xi \in \Xi \\ g \in \mathcal{H},\ g_0 \in \mathbb{R} \end{array} \right\}$$

Distributionally Robust Chance Constraints

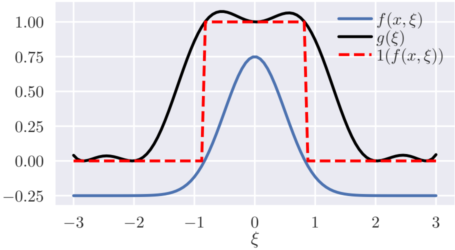
Exact reformulation

Primal formulation of feasible set:

$$Z := \{x \in \mathcal{X} : \inf_{P \in \mathcal{P}} P(f(x,\xi) \leq 0) \geq 1 - \alpha\}$$
$$= \{x \in \mathcal{X} : \sup_{P \in \mathcal{P}} \mathbb{E}_P[\mathbb{1}(f(x,\xi) \geq 0))] \leq \alpha\}$$

Assumption: Constraint function $f(x,\xi)$ is semi-continuous

Dual formulation of feasible set (Zhu u. a. 2020):



$$\mathbb{1}(f(x,\xi) > 0) \leq g(\xi) + g_0 \quad \forall \xi \in \Xi$$

- Non-convex
- semi-inf constraint

Convex conservative approximation

Using CVaR

$$\sup_{P \in \mathcal{P}} \mathrm{VaR}^{P}_{1-\alpha}(f(x,\xi)) \leq 0 \qquad \rightarrow \qquad \sup_{P \in \mathcal{P}} \mathrm{CVaR}^{P}_{1-\alpha}(f(x,\xi)) \leq 0$$

$$Z_{\mathrm{CVaR}} := \left\{ x \in \mathcal{X} : \begin{array}{l} g_0 + \frac{1}{N}\sum_{i=1}^{N} g(\xi_i) + \varepsilon \|g\|_{\mathcal{H}} \leq t\alpha \\ [f(x,\xi) + t]_+ \leq g_0 + g(\xi) \quad \forall \xi \in \Xi \\ g \in \mathcal{H}, \quad g_0, t \in \mathbb{R} \end{array} \right\}$$

$$g(\xi) = \sum_{i=1}^{N} \gamma_i k(\xi_i, \xi) \quad \rightarrow \quad ||g||_{\mathcal{H}} = \sqrt{\gamma K \gamma} \quad K_{ij} = k(\xi_i, \xi_j)$$

Convex conservative approximation

Using CVaR

$$\sup_{P\in\mathcal{P}} \mathrm{VaR}^P_{1-\alpha}(f(x,\xi)) \le 0 \quad \to \quad \sup_{P\in\mathcal{P}} \mathrm{CVaR}^P_{1-\alpha}(f(x,\xi)) \le 0$$

$$Z_{\mathrm{CVaR}} := \left\{ x \in \mathcal{X} : \begin{array}{l} g_0 + \frac{1}{N}\sum_{i=1}^{N} g(\xi_i) + \varepsilon \|g\|_{\mathcal{H}} \le t\alpha \\ [f(x,\xi)+t]_+ \le g_0 + g(\xi) \quad \forall \xi \in \Xi \\ g \in \mathcal{H}, \quad g_0, t \in \mathbb{R} \end{array} \right\}$$

$$g(\xi) = \sum_{i=1}^{N} \gamma_i k(\xi_i,\xi) \quad \to \quad ||g||_{\mathcal{H}} = \sqrt{\gamma K \gamma} \quad K_{ij} = k(\xi_i,\xi_j)$$

Constraint relaxation

$$[f(x,\xi_i)+t]_+ \le g_0 + (K\gamma)_i \quad \forall i = 1,\ldots,N$$

- Avoid ad-hoc reformulation limiting constraint function
- Statistical guarantee for constraint satisfaction

Numerical experiments

Chance constrained portfolio optimization

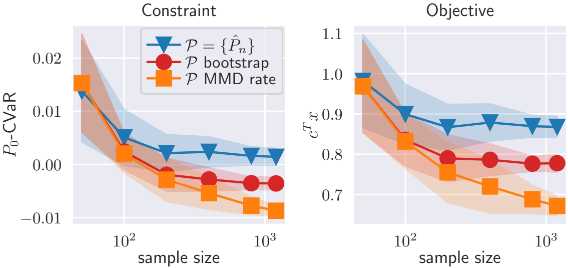
$$\max_{x \in \mathcal{X}} \quad c^T x$$

$$\text{s.t.} \quad P[(x^T \xi)^2 - 1 \leq 0] \geq 1 - \alpha$$

$$\downarrow$$

$$\sup_{P \in \mathcal{P}} \text{CVaR}^P_{1-\alpha}[(x^T \xi)^2 - 1] \leq 0$$

$$\mathcal{X} = \{x \in \mathbb{R}^3 : \sum_{i=1}^{3} x_i \leq 1, \quad x_i \geq 0 \quad \forall i = 1, 2, 3\}$$



MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

WI AS

Blanchet, Jose, Yang Kang und Karthyek Murthy (Sep. 2019). „Robust Wasserstein Profile Inference and Applications to Machine Learning“. In: *Journal of Applied Probability* 56.3, S. 830–857.

Chen, Zhi, Daniel Kuhn und Wolfram Wiesemann (Mai 2022). *Data-Driven Chance Constrained Programs over Wasserstein Balls*.

Hota, Ashish R., Ashish Cherukuri und John Lygeros (Okt. 2018). „Data-Driven Chance Constrained Optimization under Wasserstein Ambiguity Sets“. In: *arXiv:1805.06729 [cs, math]*.

Weijun, Xie (Juni 2018). „On Distributionally Robust Chance Constrained Programs With Wasserstein Distance“. In: *arXiv*.

Zhu, Jia-Jie u. a. (Dez. 2020). „Kernel Distributionally Robust Optimization“. In: *arXiv:2006.06981 [cs, math, stat]*.

CVaR reformulations

Tractable reformulation of semi-inf constraint

Ad-hoc reformulations of CVaR relaxation

- Wasserstein DR-CC: e.g., piece-wise affine constraint, quadratic constraints (Hota u. a. 2018; Weijun 2018)
  - Assumptions on support of ξ
  - Limit constraint functions
- MMD DR-CC:
  - Limit support
  - Adjust kernel choice to constraint function: e.g. linear kernel for piecewise-affine $f(x, \xi)$
    $\rightarrow$ Non-characteristic kernel limits MMD: Finite moment comparison of distributions

Reproducing Kernel Hilbert Space

Kernel DRO

Consider a positive definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$\to$ similarity measure for $x \in \mathcal{X}$

Let $\mathcal{H}$ be a Hilbert space of functions $\mathcal{X} \to \mathbb{R}$ with associated inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$

k is a reproducing kernel of $\mathcal{H}$ $\quad k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$

- $k(x, \cdot) \in \mathcal{H} \quad \forall x \in \mathcal{X}$
- $f(x) = \langle f, k(x, \cdot) \rangle_{\mathcal{H}} \quad \forall f \in \mathcal{H}, x \in \mathcal{X}$

Kernel Mean Embedding:

$\mu_P(\cdot) = \int_{\mathcal{X}} k(x, \cdot) dP(x) \qquad \mathbb{E}[f(x)] = \int_{\mathcal{X}} f(x) dP(x) = \langle f, \mu_p \rangle_{\mathcal{H}}$

$$\mathrm{MMD}(P, Q) := ||\mu_P - \mu_Q||_{\mathcal{H}}$$

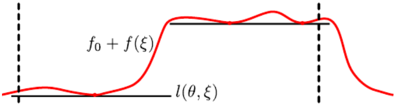
Kernel DRO(Zhu u. a. 2020)

An introduction

- Primal:
  $\min_\theta \sup_{P \in K} \int \ell(\theta, \xi) dP(\xi)$
- Dual:

$$\min_{\theta, f_0, \alpha} \quad f_0 + \delta_K(f)$$
$$\text{subject to} \quad \ell(\theta, \xi) \le f_0 + f(\xi), \forall \xi \in \mathcal{X}$$

with
$f(\xi) = \sum_i \alpha_i \cdot k(\xi_i, \xi)$,
$f(\xi) \in \mathcal{H}, f_0 \in \mathbb{R}$.



RKHS-function f such that it majorizes ℓ

Finite-sample guarantee

Concentration rate of MMD:

$$\mathrm{MMD}(P_{true}, \hat{P}) \leq \sqrt{\frac{C}{N}} + \sqrt{\frac{2C\log(1/\delta)}{N}},$$

with probability $1-\delta$ and $C \geq \sup_x k(x,x)$

$$\mathrm{CVaR}^{P}_{1-\alpha}[f(x,\xi)] \leq M\sqrt{\frac{2\log(1/\delta)}{N}}$$

Finite-sample guarantee

Concentration rate of MMD:

$$\mathrm{MMD}(P_{true}, \hat{P}) \leq \sqrt{\frac{C}{N}} + \sqrt{\frac{2C \log(1/\delta)}{N}},$$

with probability $1-\delta$ and $C \geq \sup_x k(x,x)$

$$\mathrm{CVaR}^{P}_{1-\alpha}[f(x,\xi)] \leq M\sqrt{\frac{2\log(1/\delta)}{N}}$$

Chance constraint guarantee

$$P(f(x,\xi) \leq M\sqrt{\frac{2\log(1/\delta)}{N}}) \geq 1-\alpha,$$

with probability $1-\delta$.

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

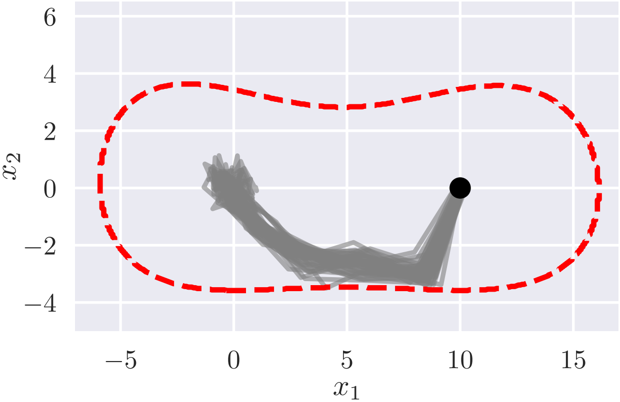
WI AS

Numerical experiments

MPC with nonlinear constraint

Linear tube-based MPC

- Nonlinear SVM constraint
- Linear dynamics



MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS